BMS

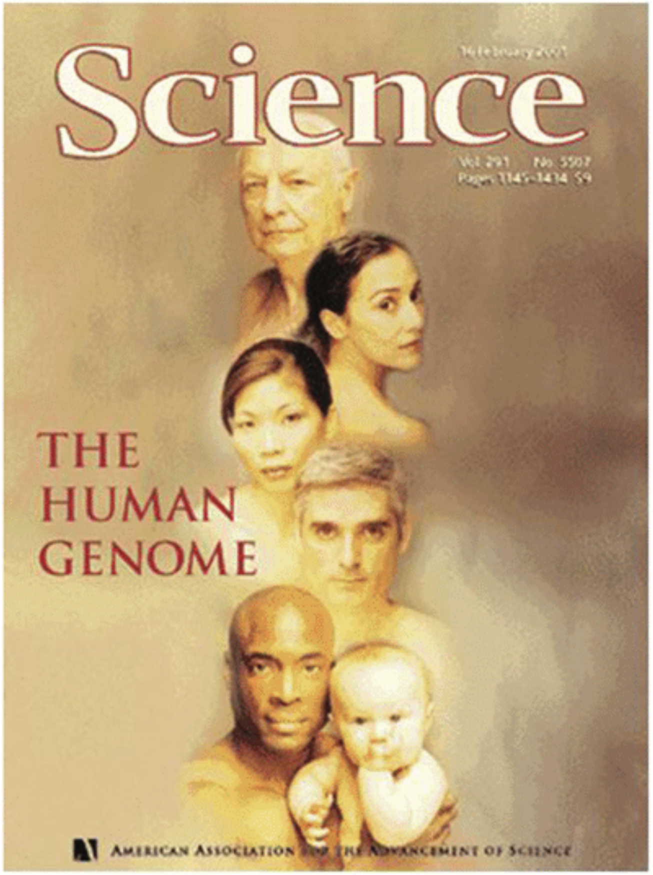
Next generation sequencing

생물정보팀 전형린

Contents

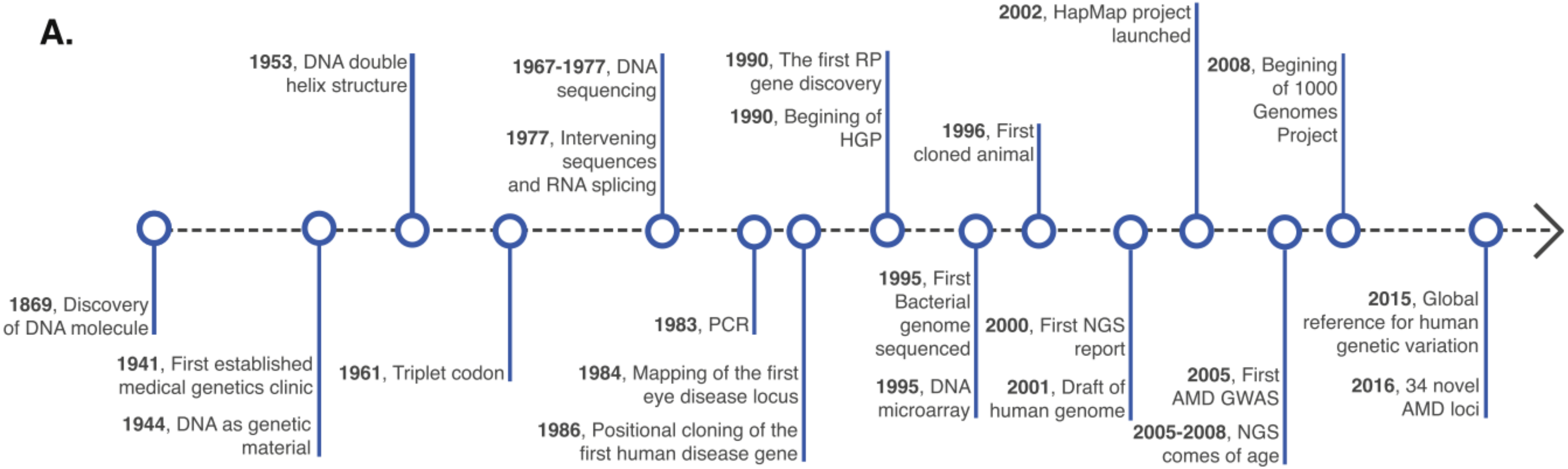
- **1주차 Next generation sequencing 및 bioinformatics 개요**
  - High throughtput DNA decode
  - 세대 별 시퀀싱 차이
  - Next generation sequencing 원리
  - Application based on NGS
- **2주차 Hybrid-Seq 개요 및 실습**
  - Hybrid-Seq 원리
  - Hybrid-Seq 분석 과정 및 결과 설명
  - 분석 실습
- **3주차 R을 이용한 plotting**
  - ggplot2 을 이용한 plotting
  - R을 이용한 phylogenetic tree plotting

Human Genome Project



https://www.researchgate.net/figure/Human-genome-Project-published-in-Nature-http-wwwnaturecom-nature-journal_fig2_329830959

Timeline for Human Genome Project



A.

1869, Discovery of DNA molecule

1941, First established medical genetics clinic

1944, DNA as genetic material

1953, DNA double helix structure

1961, Triplet codon

1967-1977, DNA sequencing

1977, Intervening sequences and RNA splicing

1983, PCR

1984, Mapping of the first eye disease locus

1986, Positional cloning of the first human disease gene

1990, The first RP gene discovery

1990, Begining of HGP

1995, First Bacterial genome sequenced

1995, DNA microarray

1996, First cloned animal

2000, First NGS report

2001, Draft of human genome

2002, HapMap project launched

2005, First AMD GWAS

2005-2008, NGS comes of age

2008, Begining of 1000 Genomes Project

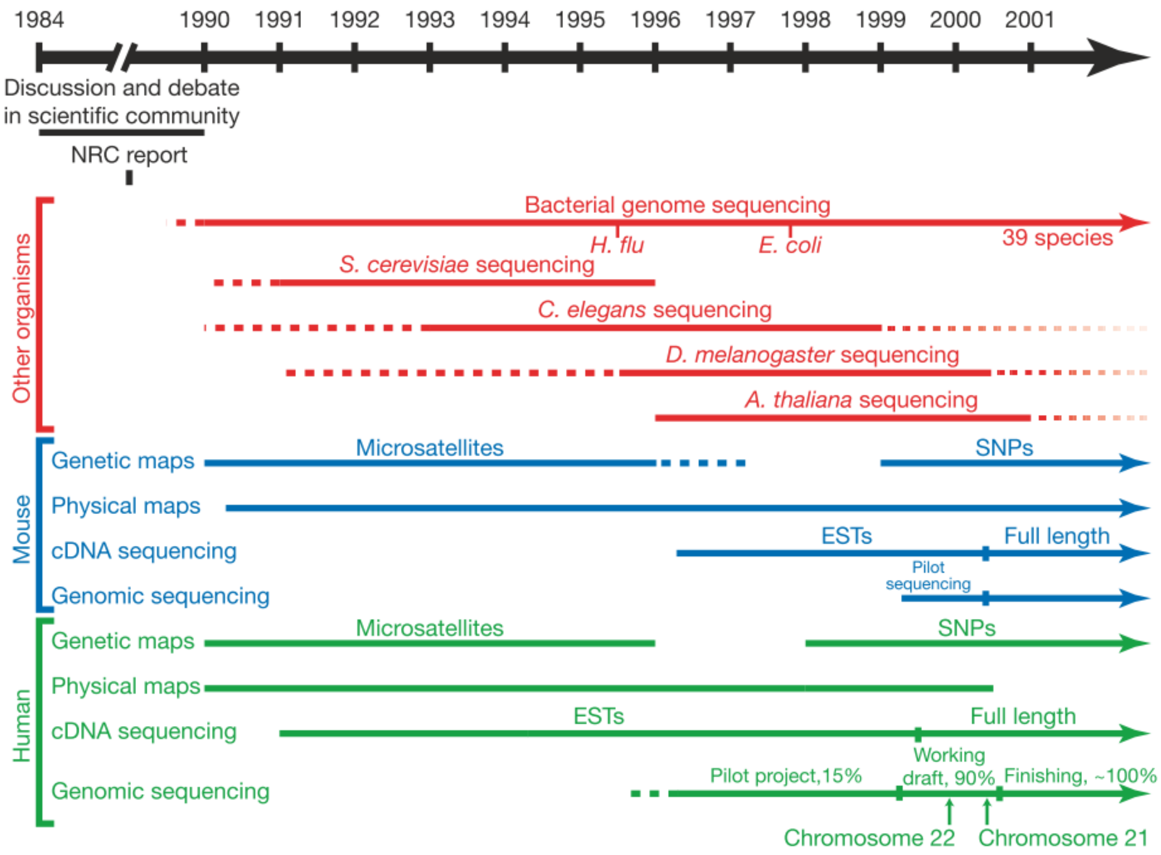
2015, Global reference for human genetic variation

2016, 34 novel AMD loci

https://www.oreilly.com/ideas/genomics-and-the-role-of-big-data-in-personalizing-the-healthcare-experience

Timeline until HGP

- **Human genome project formally launched in 1990 and was declared complete on April 14, 2003**

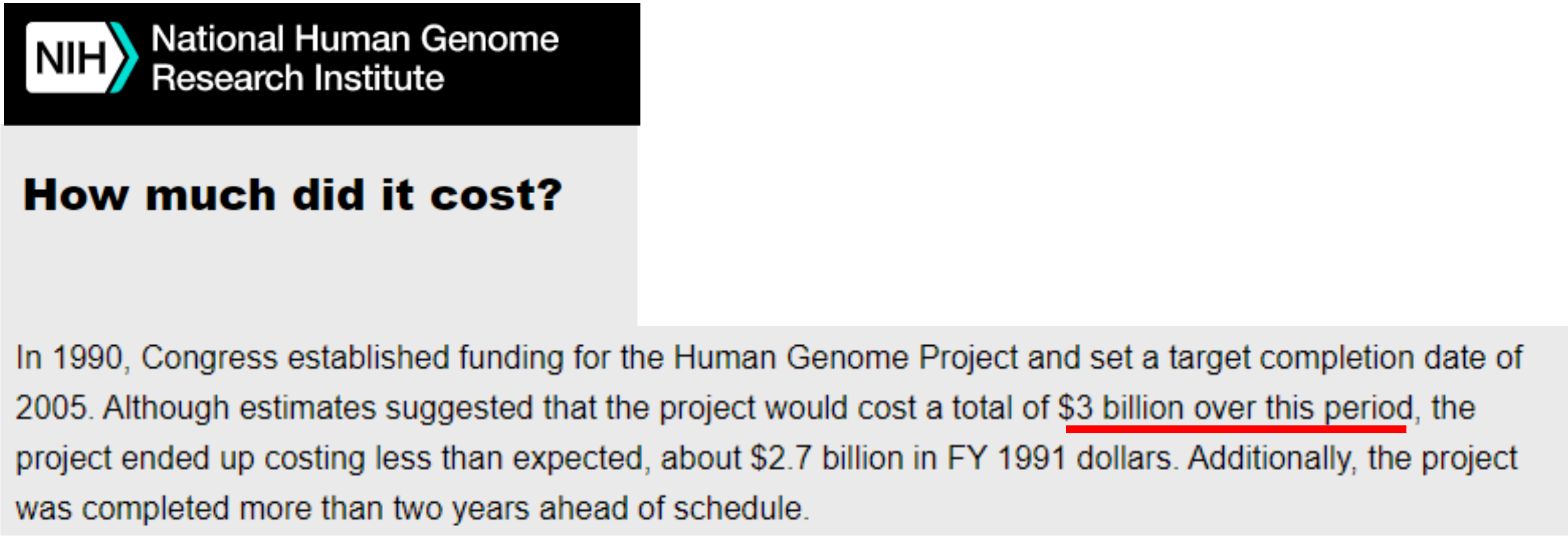


Nature, 2001

White House on 26 June 2000

https://www.theguardian.com/science/gallery/2010/jun/23/human-genome-project

Cost for Human Genome Project



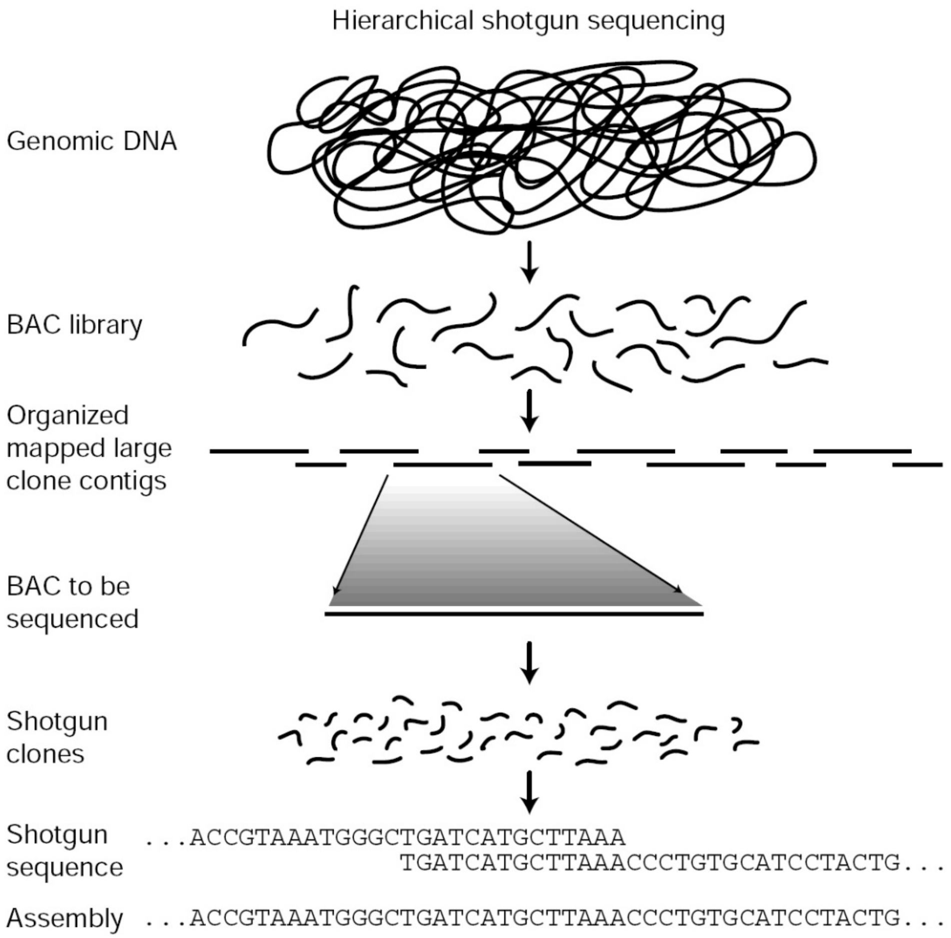
NIH National Human Genome Research Institute

How much did it cost?

In 1990, Congress established funding for the Human Genome Project and set a target completion date of 2005. Although estimates suggested that the project would cost a total of $3 billion over this period, the project ended up costing less than expected, about $2.7 billion in FY 1991 dollars. Additionally, the project was completed more than two years ahead of schedule.

https://www.researchgate.net/figure/Human-genome-Project-published-in-Nature-http-wwwnaturecom-nature-journal_fig2_329830959

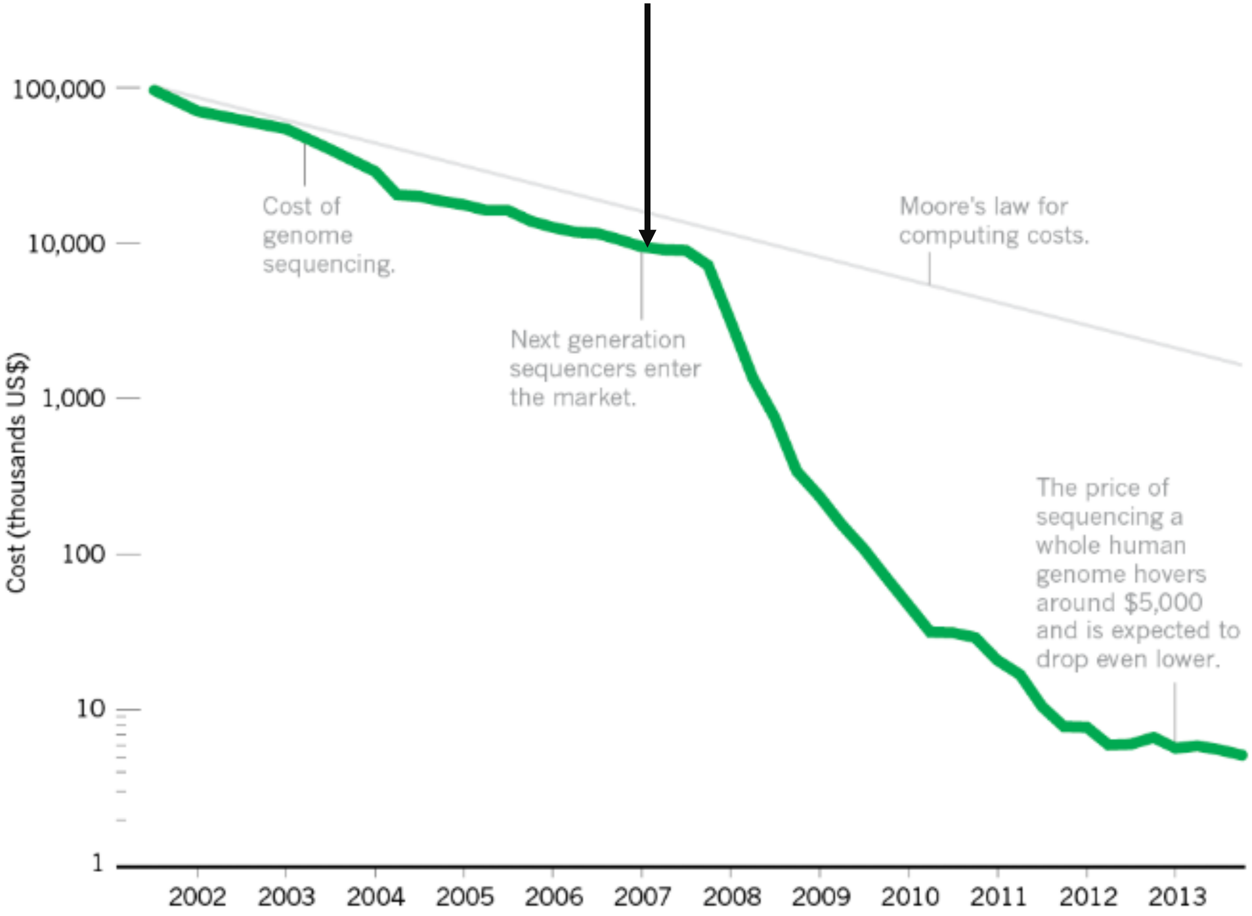
Strategy for sequencing


Hierarchical shotgun sequencing
Genomic DNA
BAC library
Organized mapped large clone contigs
BAC to be sequenced
Shotgun clones
Shotgun sequence
...ACCGTAAATGGGCTGATCATGCTTAAA
TGATCATGCTTAAACCCTGTGCATCCTACTG...
Assembly
...ACCGTAAATGGGCTGATCATGCTTAAACCCTGTGCATCCTACTG...

Nature, 2001

Cost for genome sequencing

Next Generation Sequencing platform was appeared as game exchanger in sequencing field



https://www.researchgate.net/figure/a-Cost-per-base-of-the-different-sequencing-techniques-as-a-function-of-time-The-gray_fig2_271772842

Next Generation Sequencing Platform

- Roche/454 (GS FLX+/GS Junior)
- Illumina Genome Analyzer (HiSeq/MiSeq/NextSeq)
- Life Technologies (3500 Genetic Analyzer)
- Ion Torrent Proton/PGM)
- Applied Biosystems (SOLiD, 3730xl DNA Analyzer)

illumina

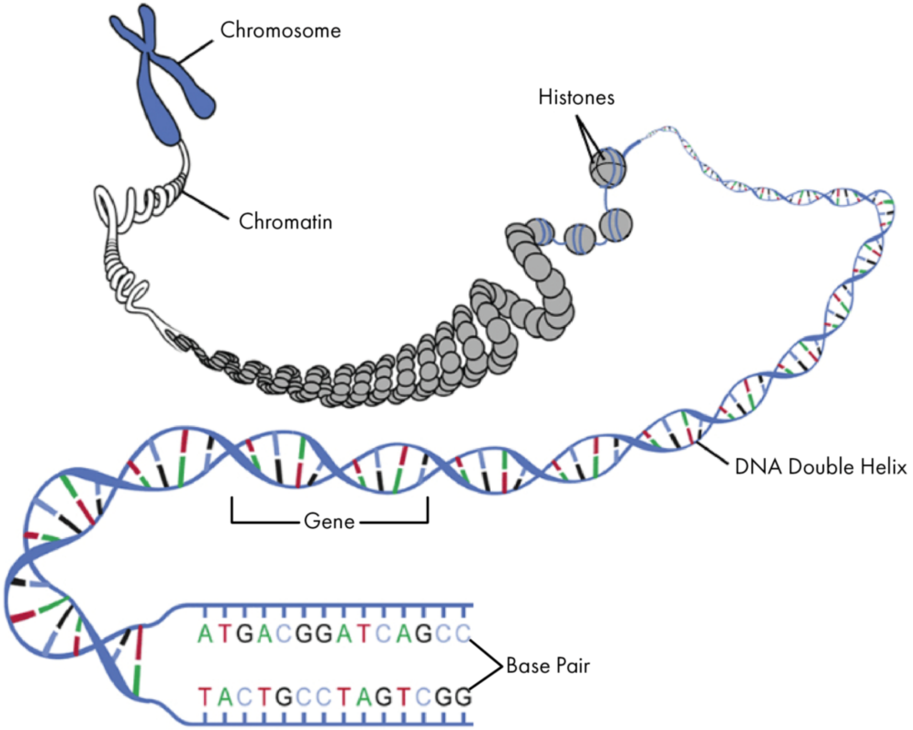
ion torrent

Roche

http://www.vib.be/en/about-vib/annual-report/2012/research/activities/Pages/Service%20Facilities.aspx

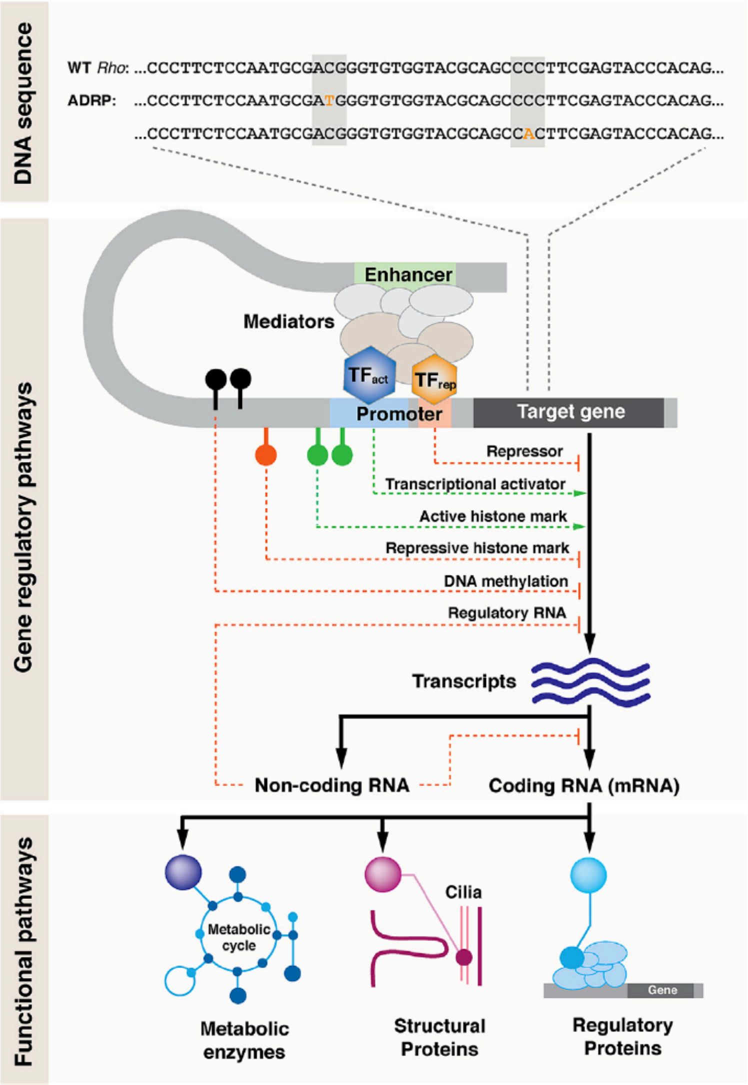
“DNA Sequencing”

- DNA sequencing involves the use of various methods for determining the order of the nucleotide bases — **adenine**, **cytosine**, **guanine**, and **thymine** — in a molecule of DNA



https://www.mygenefood.com/the-language-of-genetics/

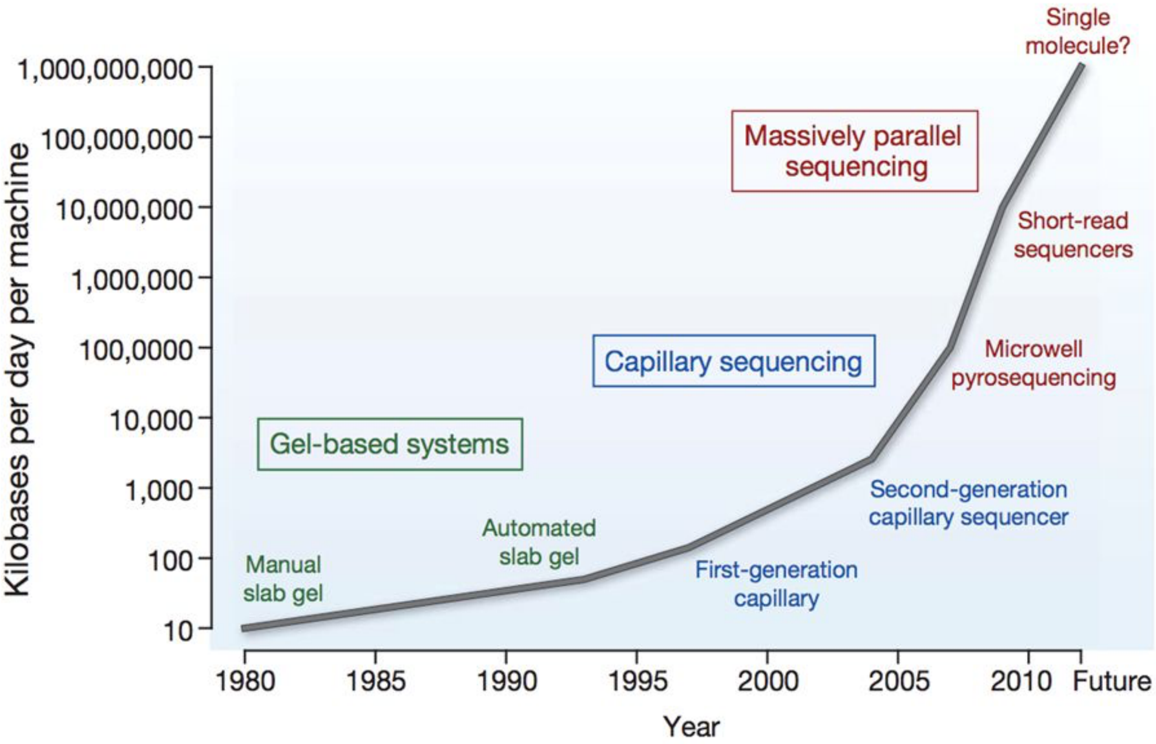
Information including in sequence



- DNA sequence includes not only template for transcription, but also act as regulation factors

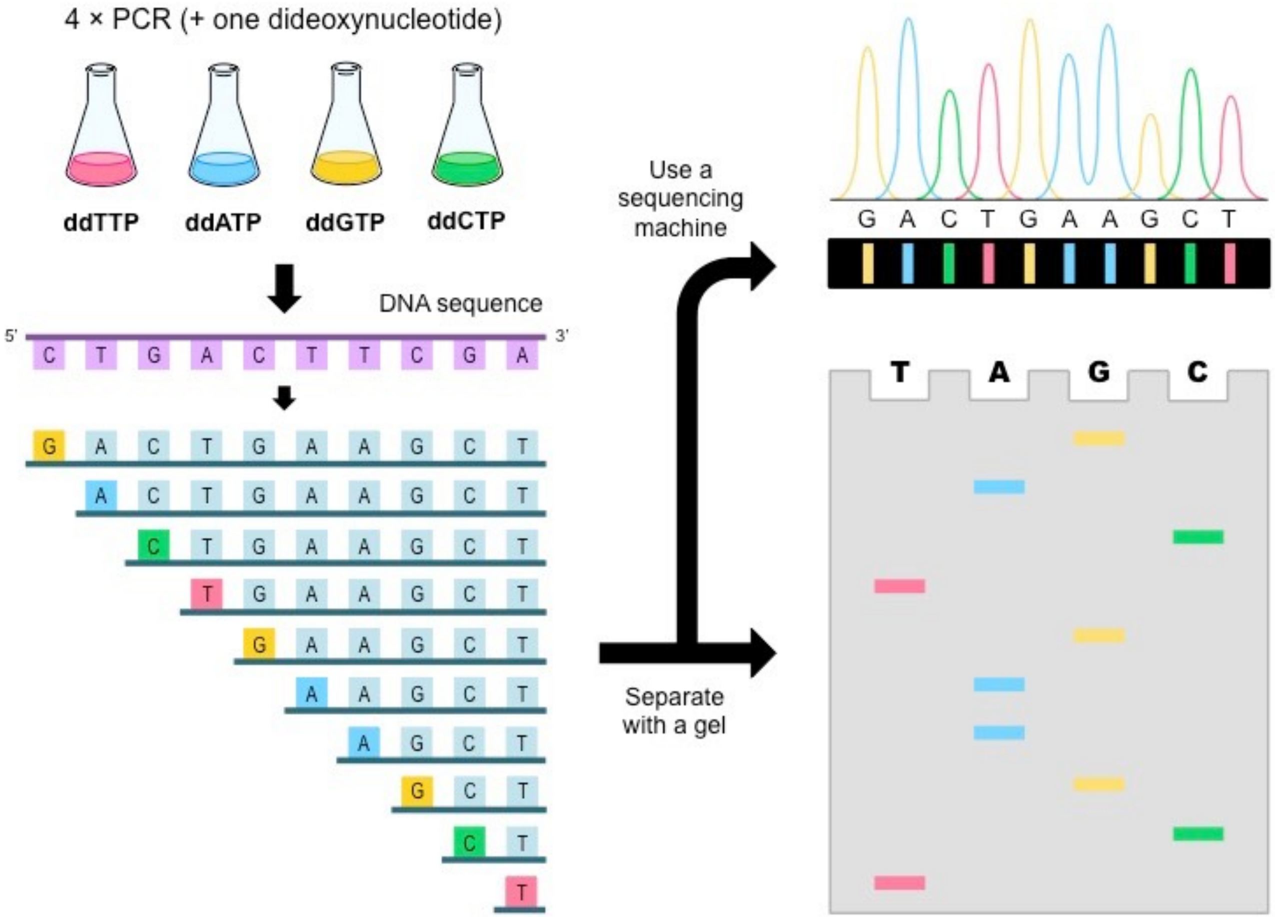
Progress in Retinal and Eye Research, 2015

History of Sequencing technology



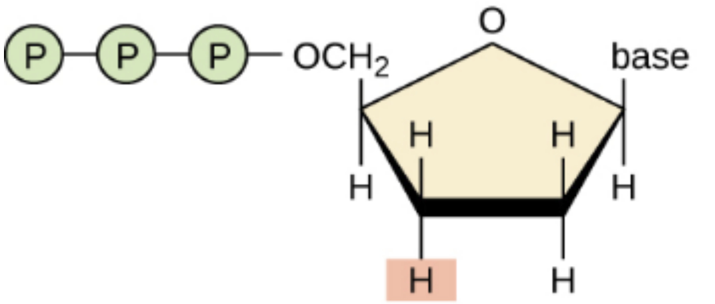
https://slideplayer.com/slide/3351231/

Sanger Sequencing

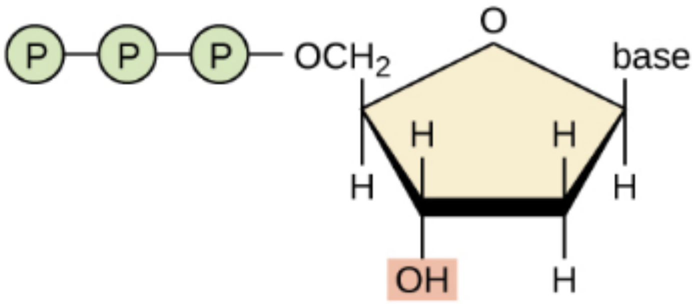


https://alevelbiology.co.uk/notes/human-genome-project/

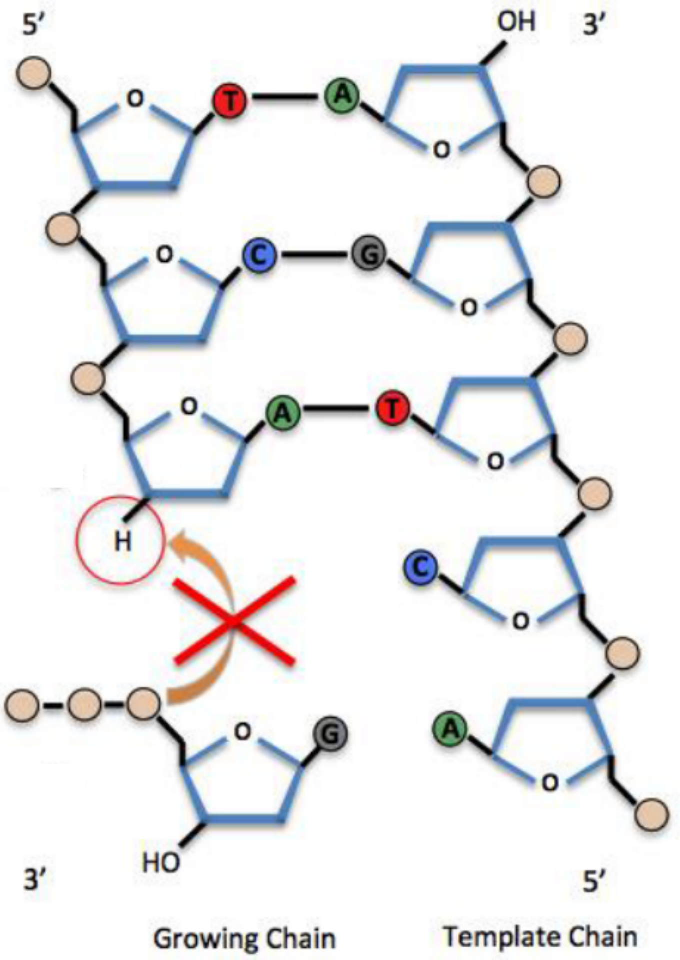
Sanger Sequencing



dideoxynucleotide (ddNTP)



deoxynucleotide (dNTP)



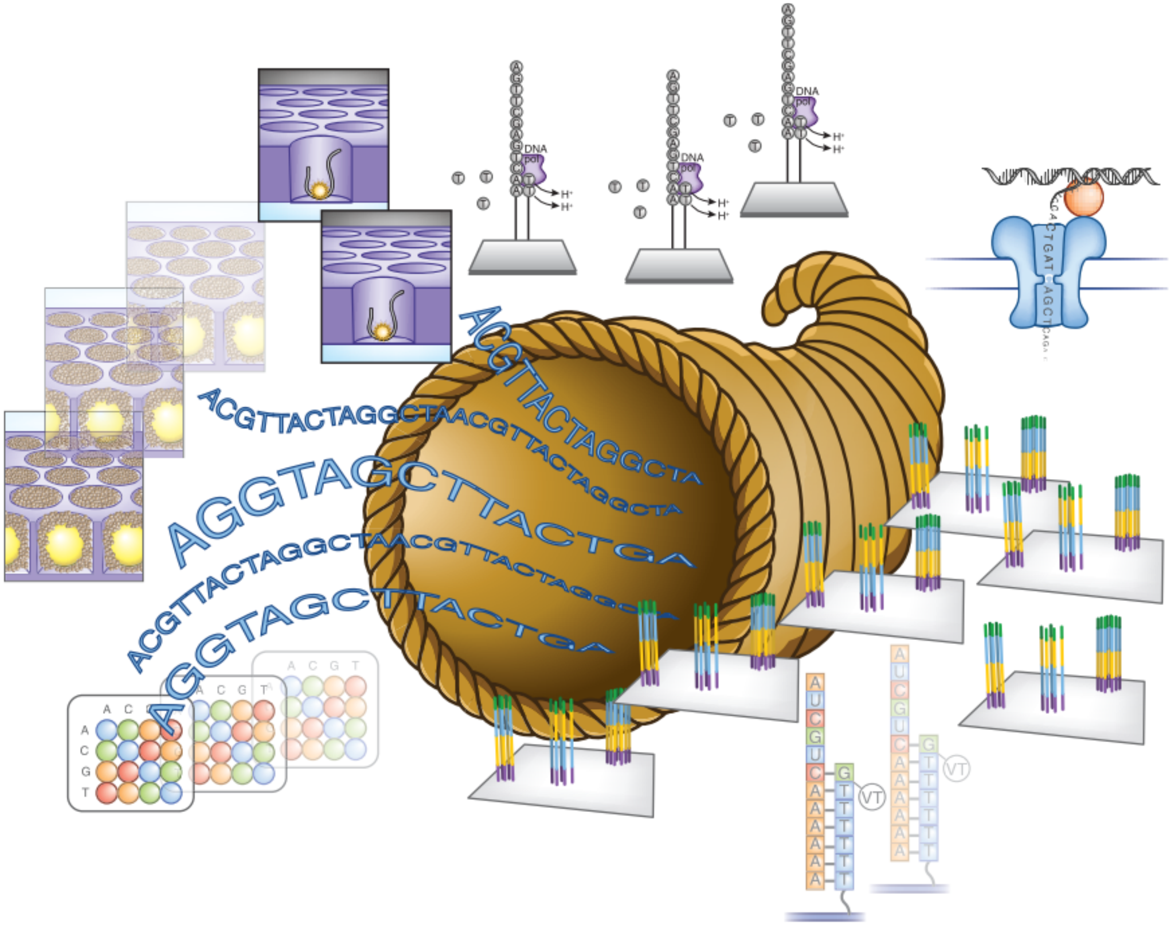
https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0166738

Sanger Sequencing

Technology	Analysis time	Average read length	Throughput (Mb/h)
Slab gel	6–8 hours	700 bp	0.0672
Capillary array electrophoresis	1–3 hours	700 bp	0.166

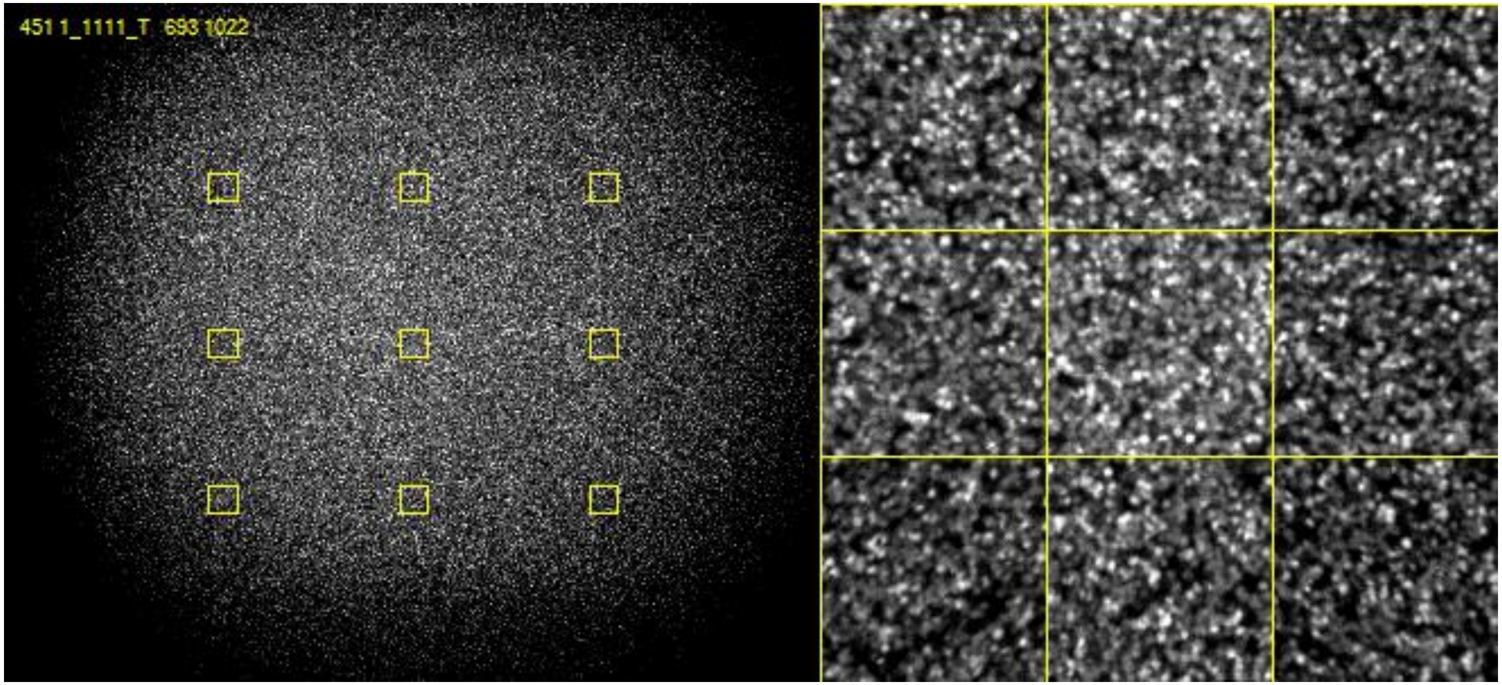
https://en.wikipedia.org/wiki/Sanger_sequencing

Next generation Sequencing



***Nature Methods*, 2014**

Next generation Sequencing



Definition of “Massive parallel”

- **Massive**

  [형용사] (육중하면서) 거대한

- **Parallel**

  [형용사] **병행[병렬]의**

- **Massive parallel sequencing**

  The DNA is sequenced via spatially separated, clonally amplified DNA templates or single DNA molecules in a flow cell.

Massive parallel sequencing = Next Generation Sequencing

Next Generation Sequencing Platform

- Roche/454 (GS FLX+/GS Junior)
- Illumina Genome Analyzer (HiSeq/MiSeq/NextSeq)
- Life Technologies (3500 Genetic Analyzer)
- Ion Torrent Proton/PGM)
- Applied Biosystems (SOLiD, 3730xl DNA Analyzer)

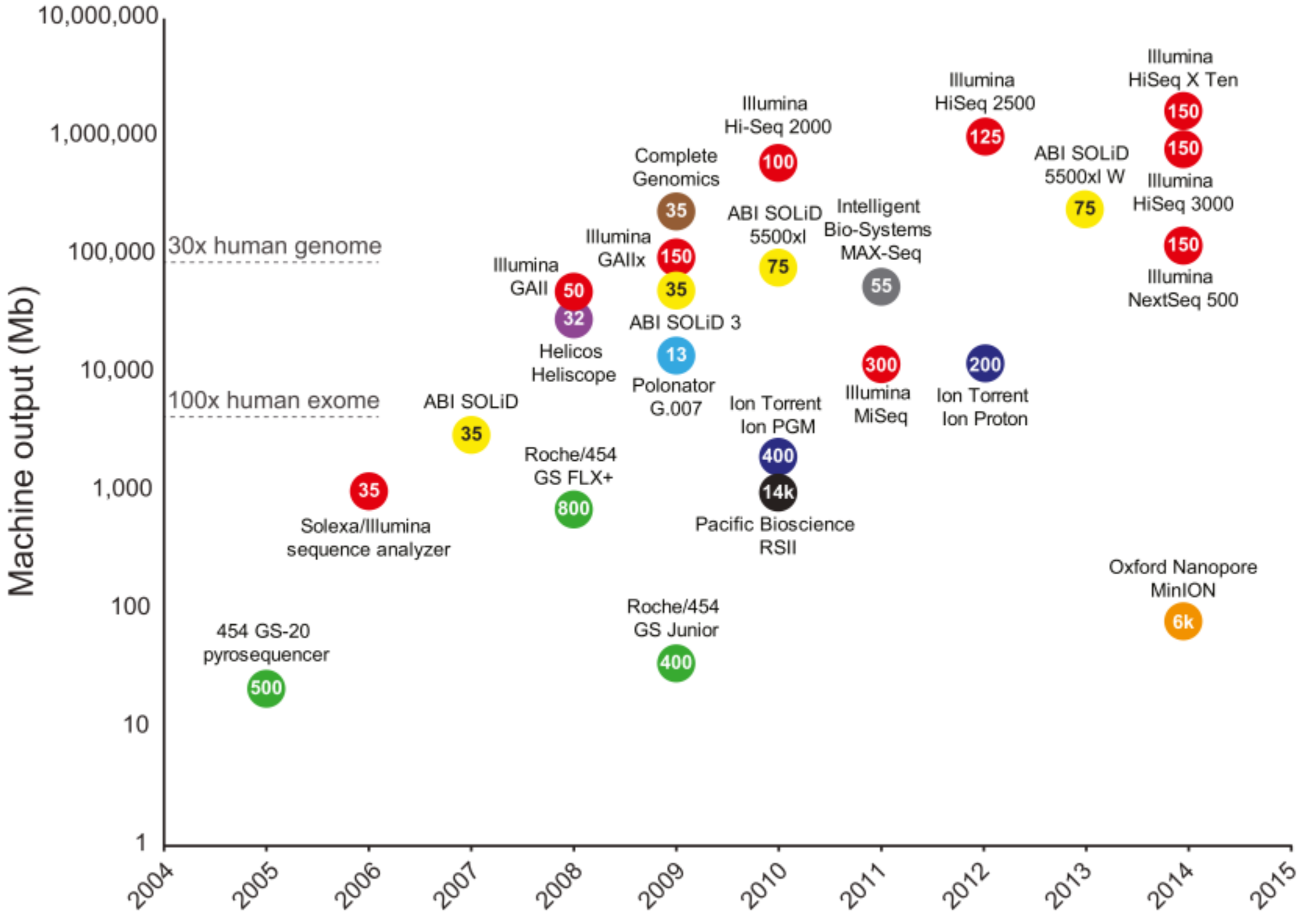
illumina

ion torrent

Roche

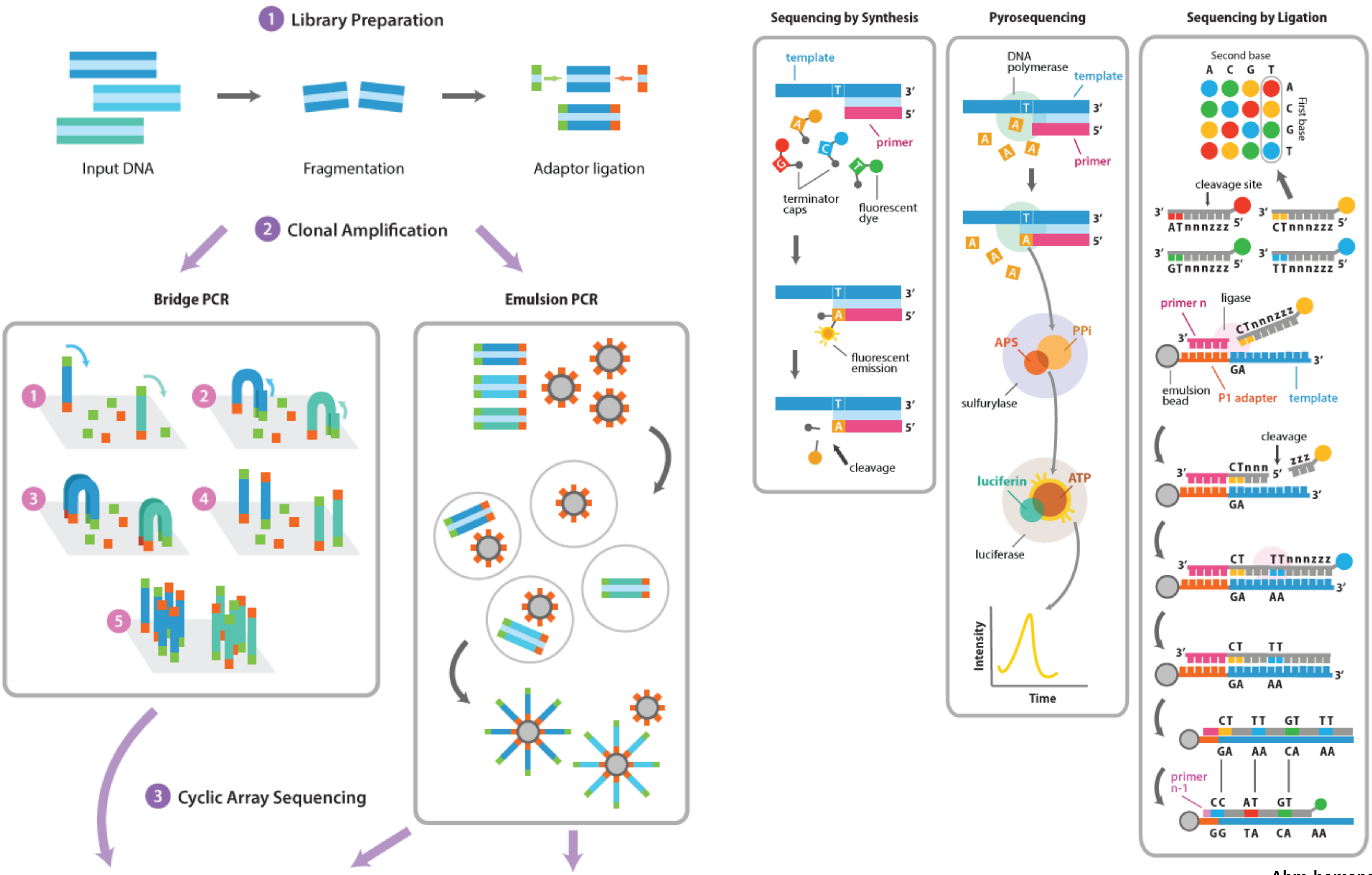
http://www.vib.be/en/about-vib/annual-report/2012/research/activities/Pages/Service%20Facilities.aspx

Next Generation Sequencing Machine



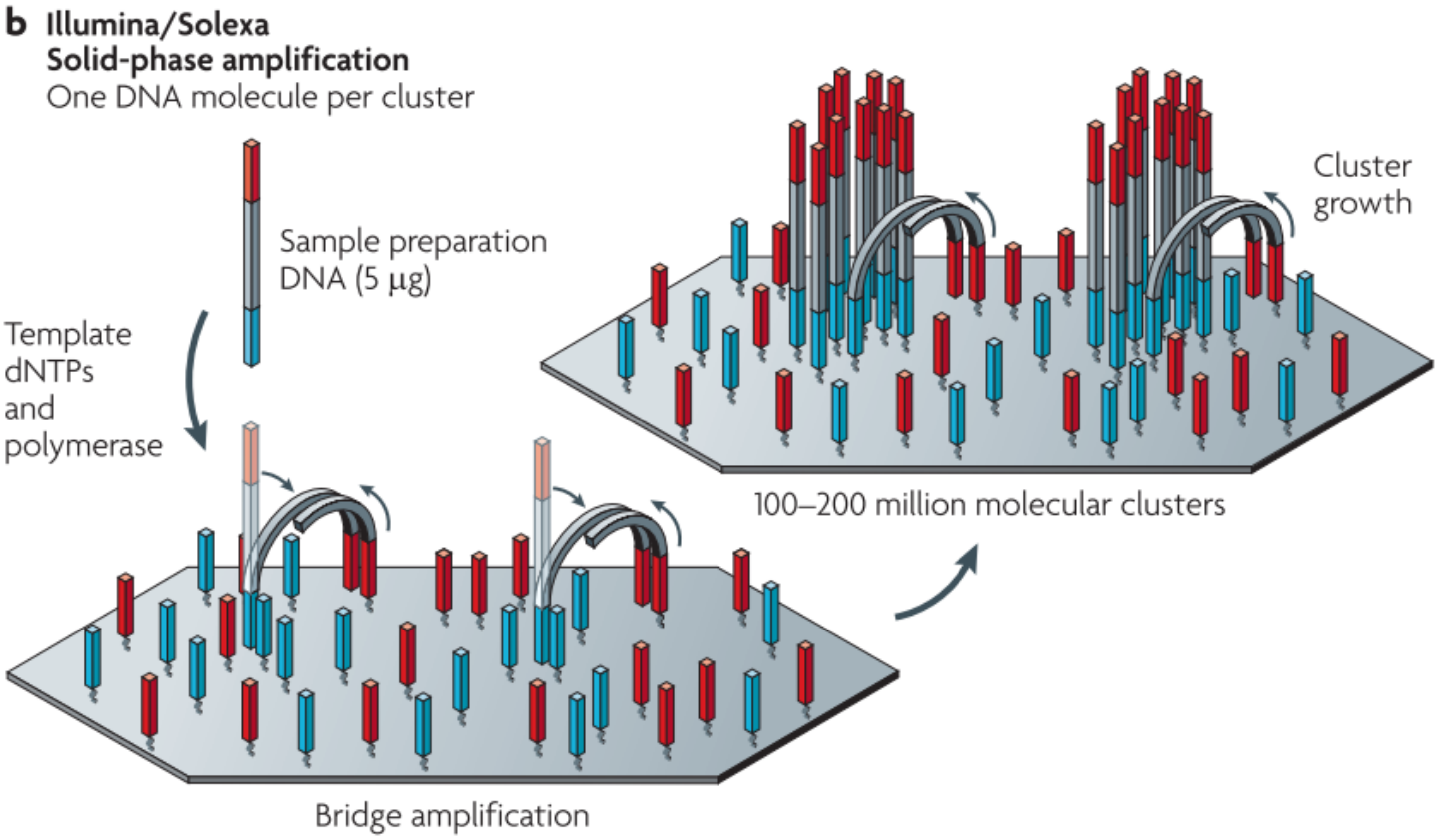
Molecular cell, 2015

Steps of NGS



Abm homepage

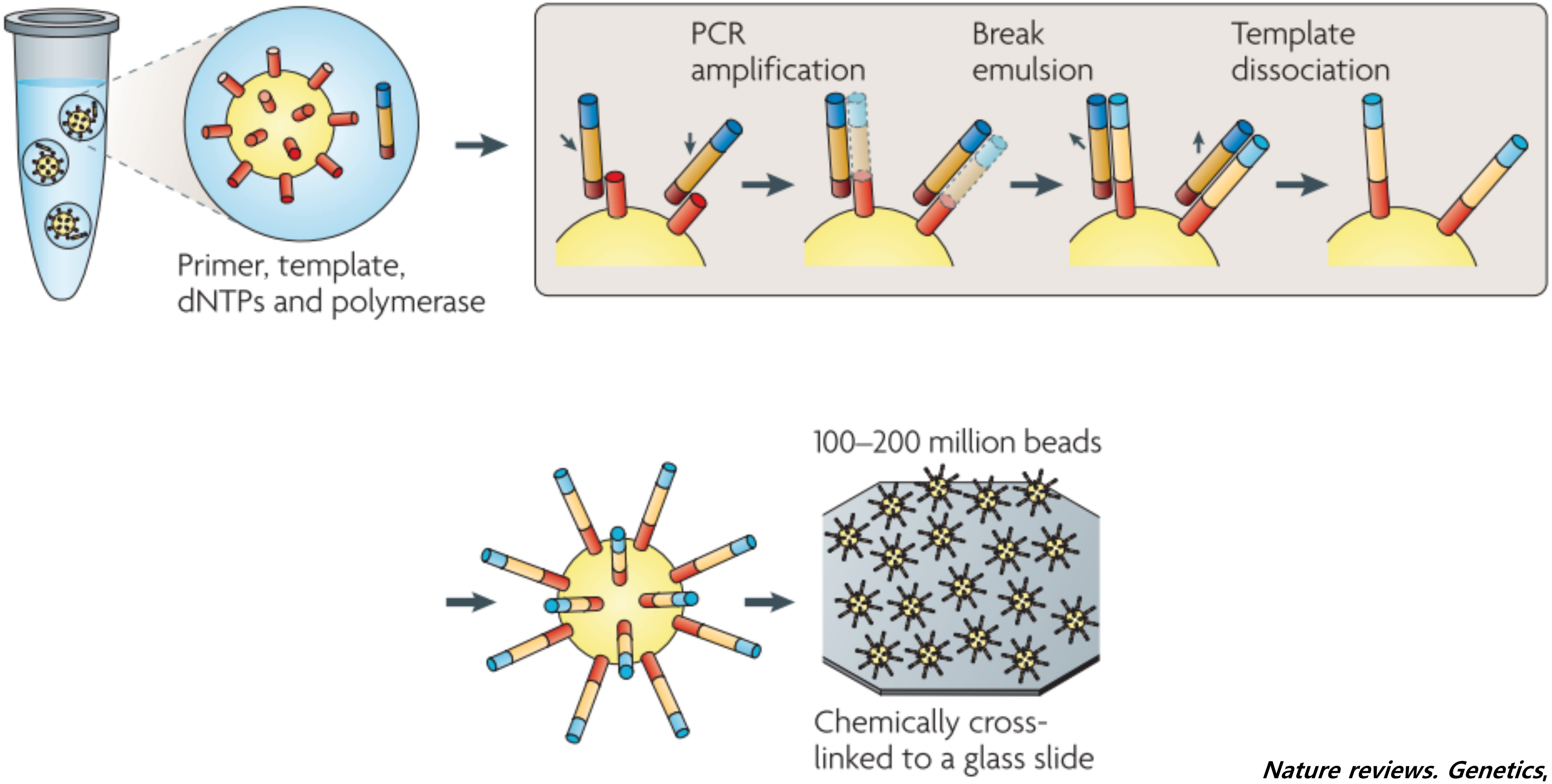
Clonal amplification - Bridge PCR



Nature reviews. Genetics, 2009

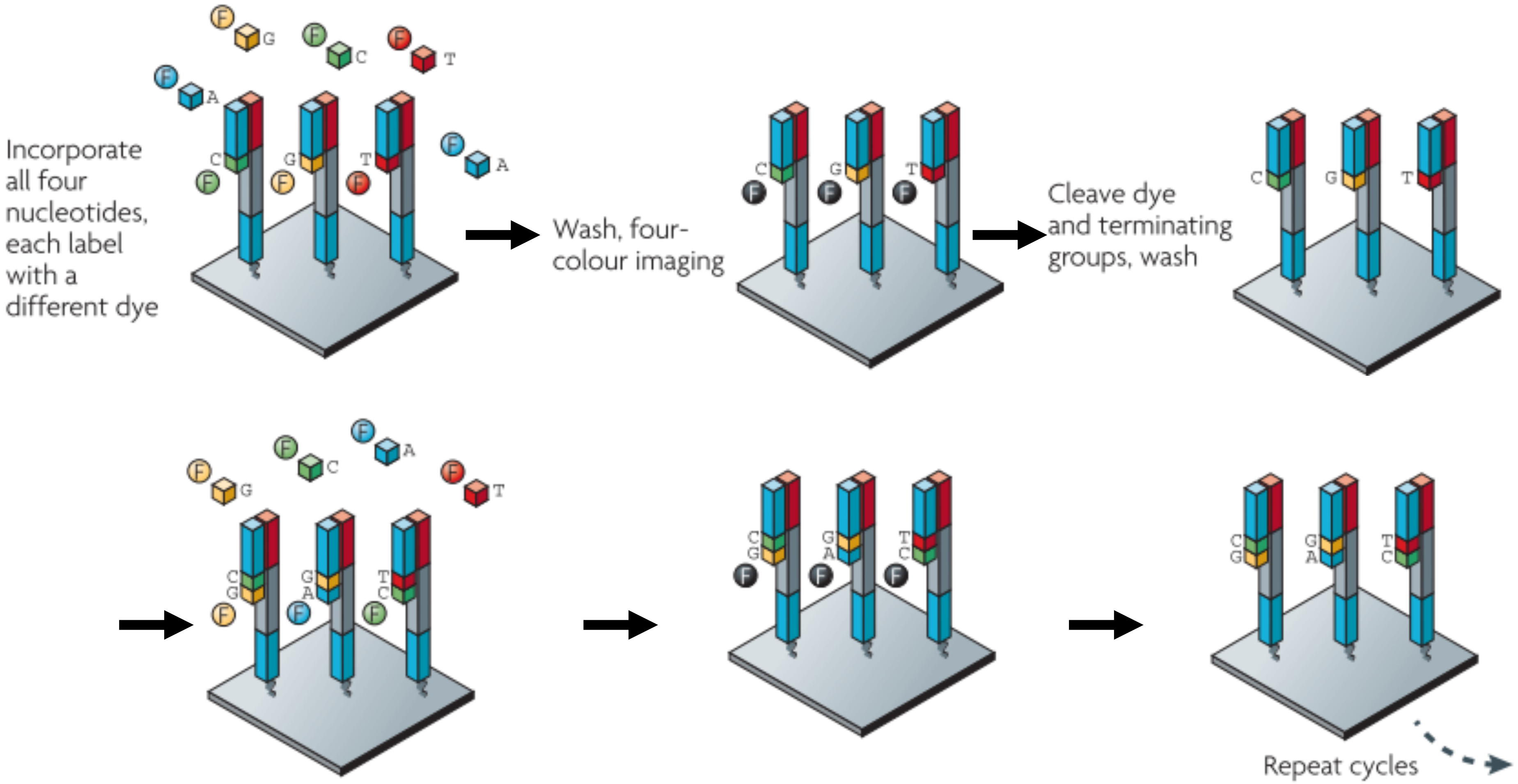
Clonal amplification – Emulsion PCR

a Roche/454, Life/APG, Polonator
Emulsion PCR
One DNA molecule per bead. Clonal amplification to thousands of copies occurs in microreactors in an emulsion



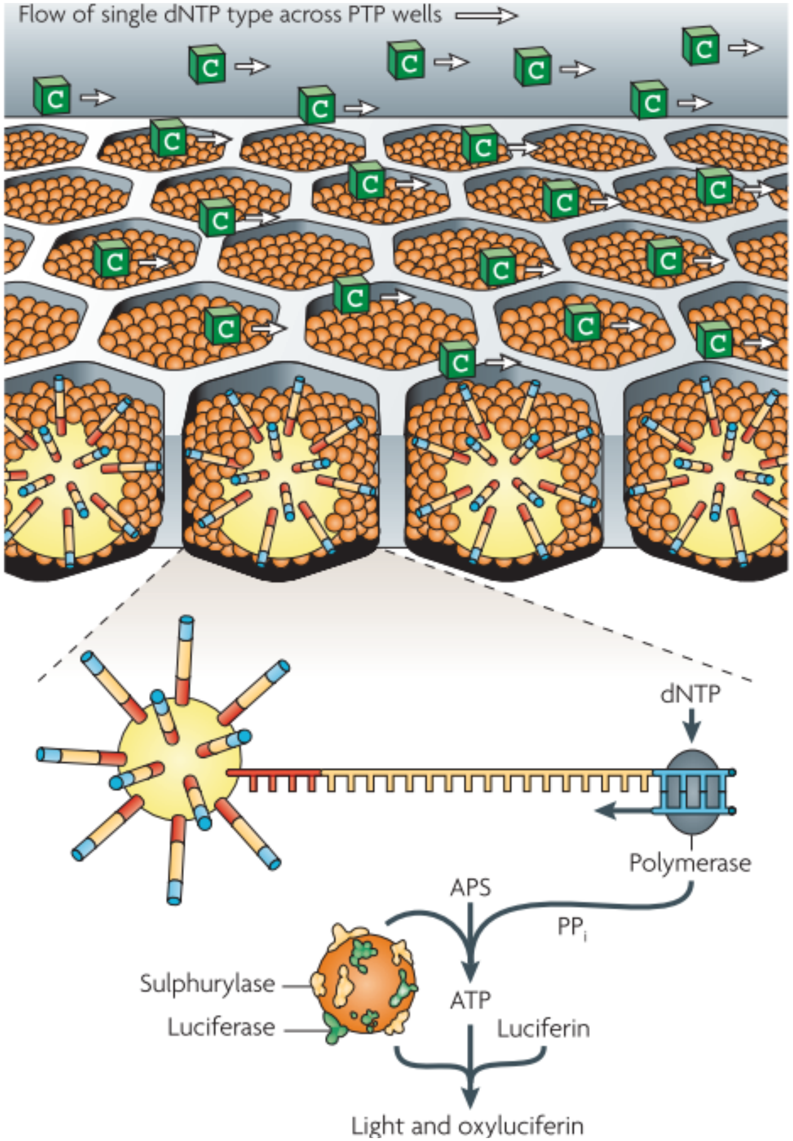
Nature reviews. Genetics, 2009

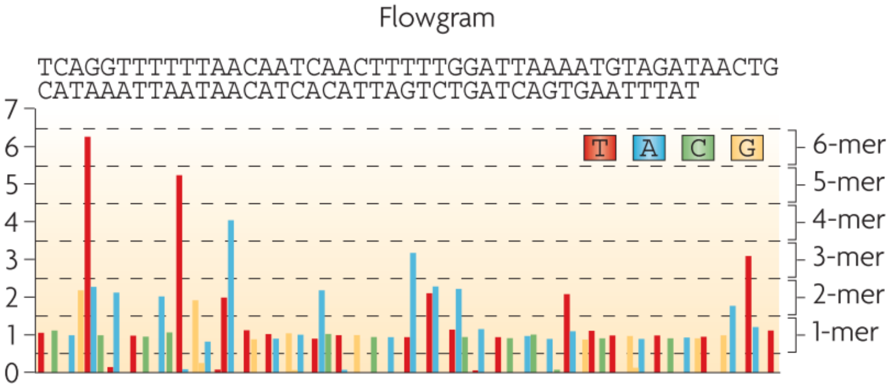
Sequencing by sequencing


Incorporate all four nucleotides, each label with a different dye
Wash, four-colour imaging
Cleave dye and terminating groups, wash
Repeat cycles
F
A
C
G
T

***Nature reviews. Genetics*, 2009**

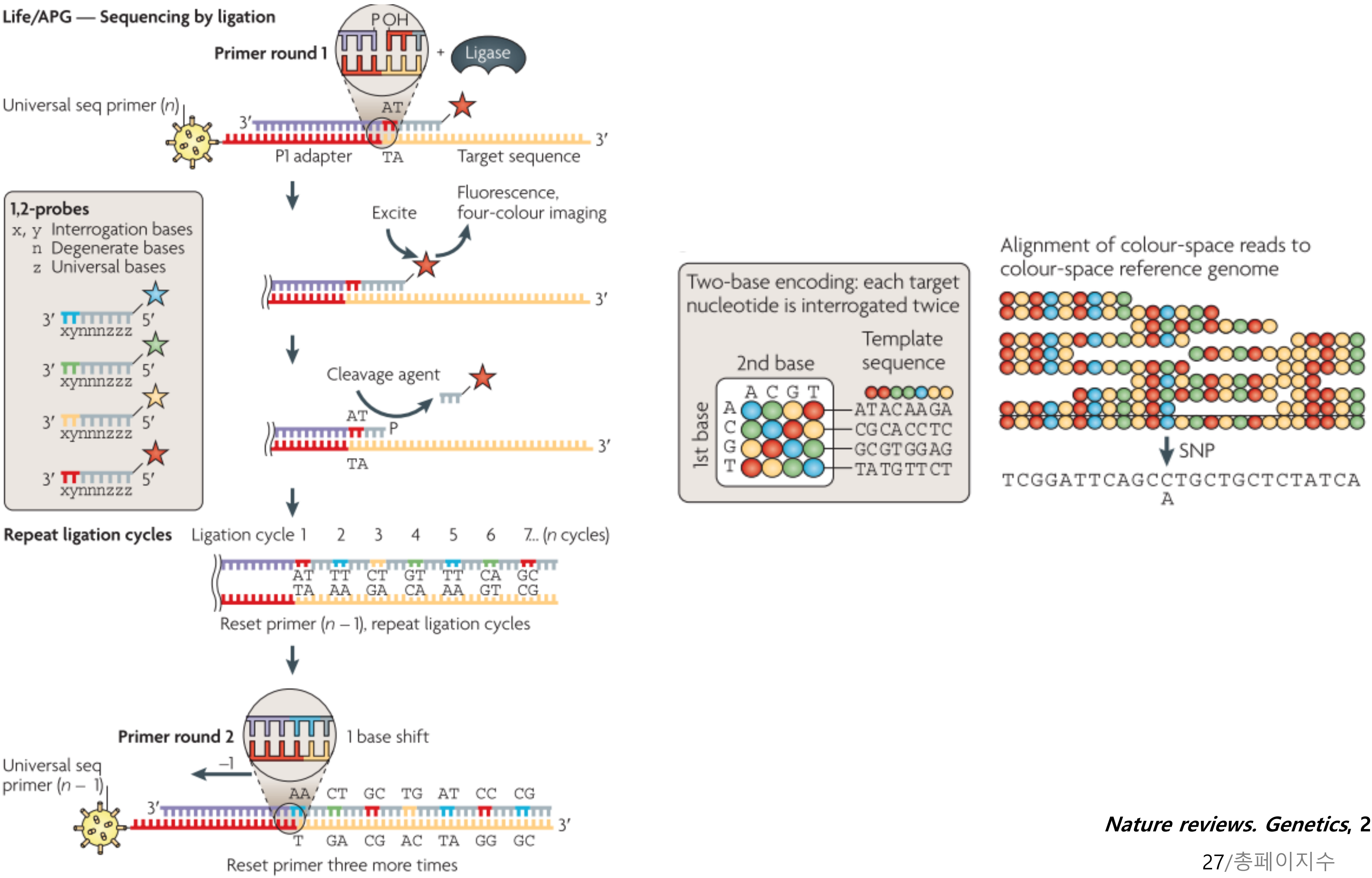
Pyrosequencing





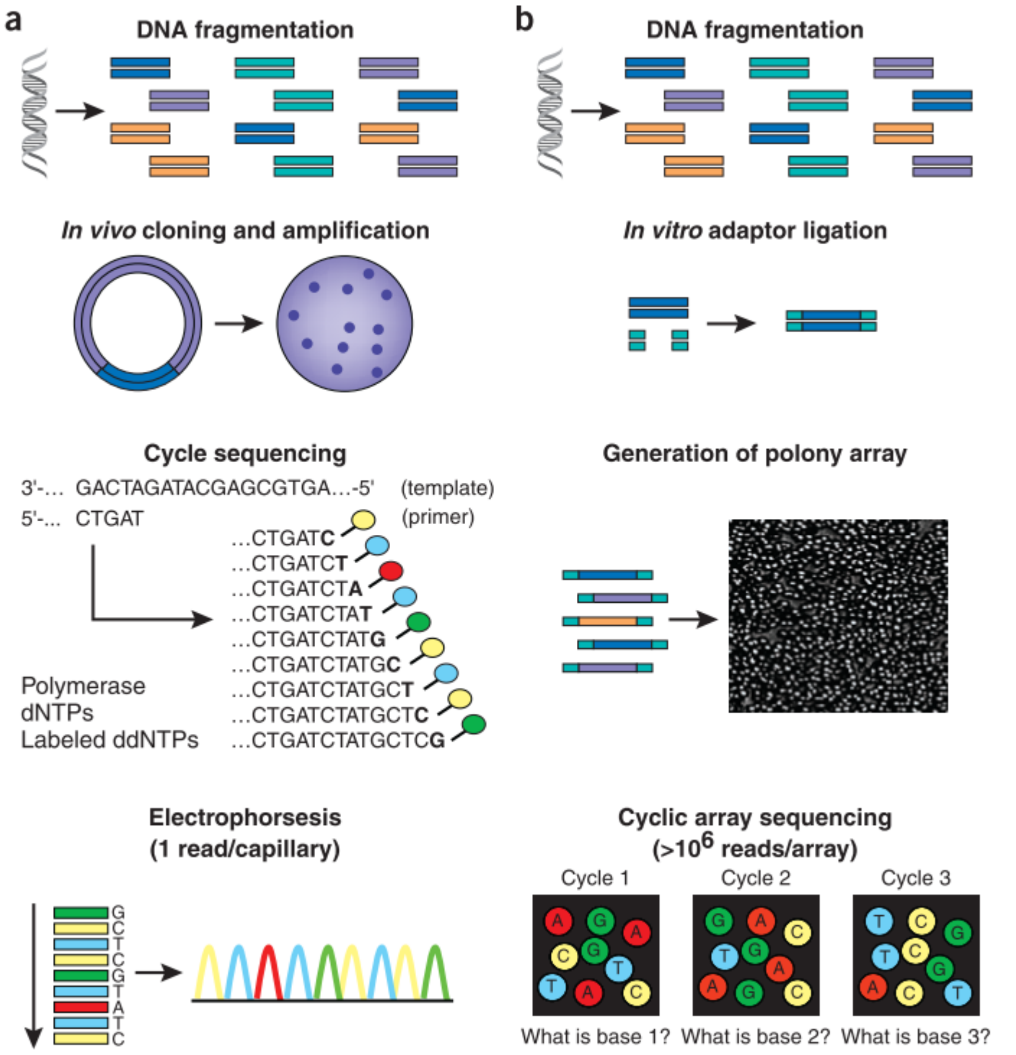
***Nature reviews. Genetics*, 2009**

Sequencing by ligation



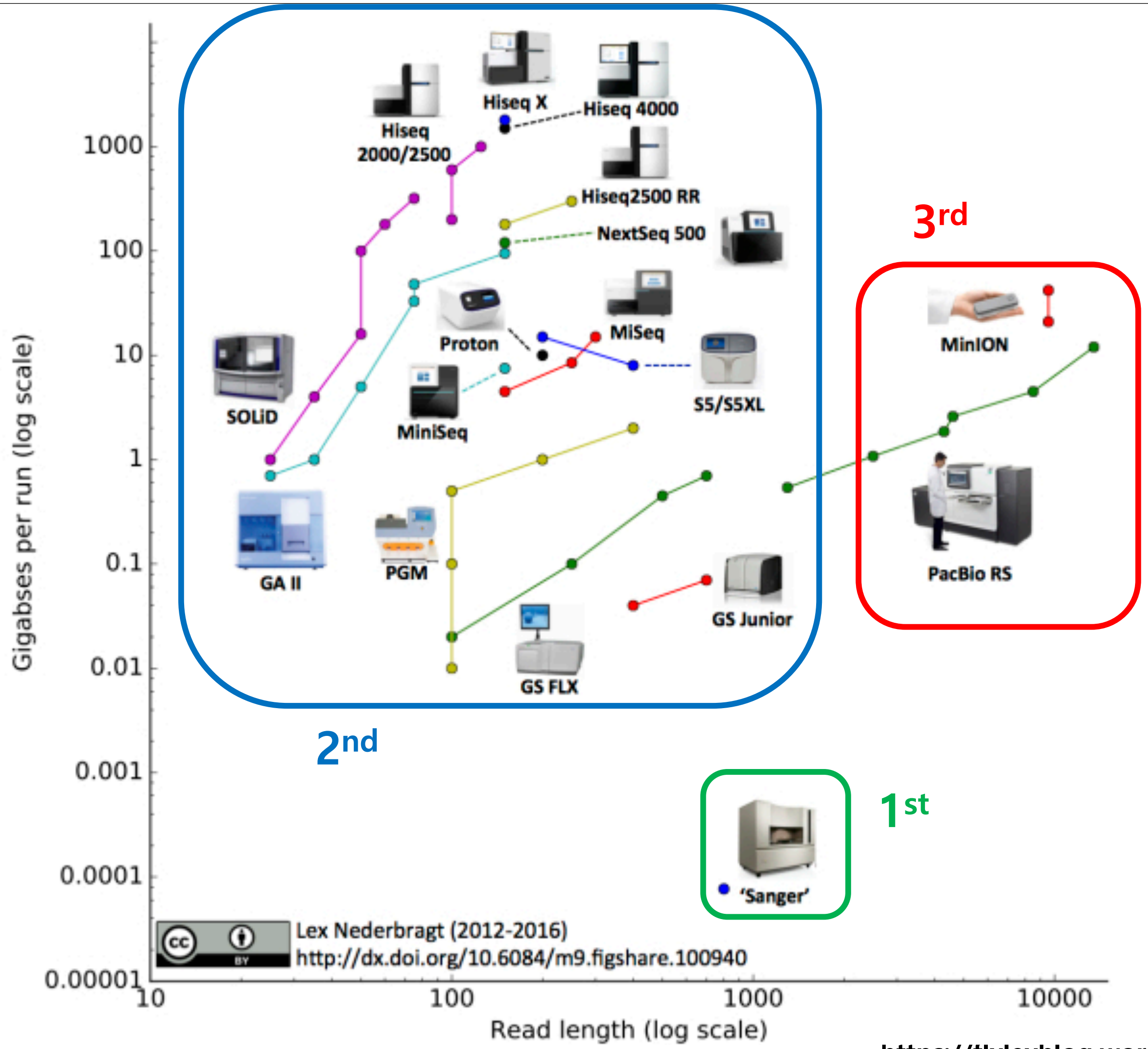
Nature reviews. Genetics, 2009

Comparison between Sanger and NGS



Nature biotechnology, 2008

History of Sequencing technology



Lex Nederbragt (2012-2016)
http://dx.doi.org/10.6084/m9.figshare.100940

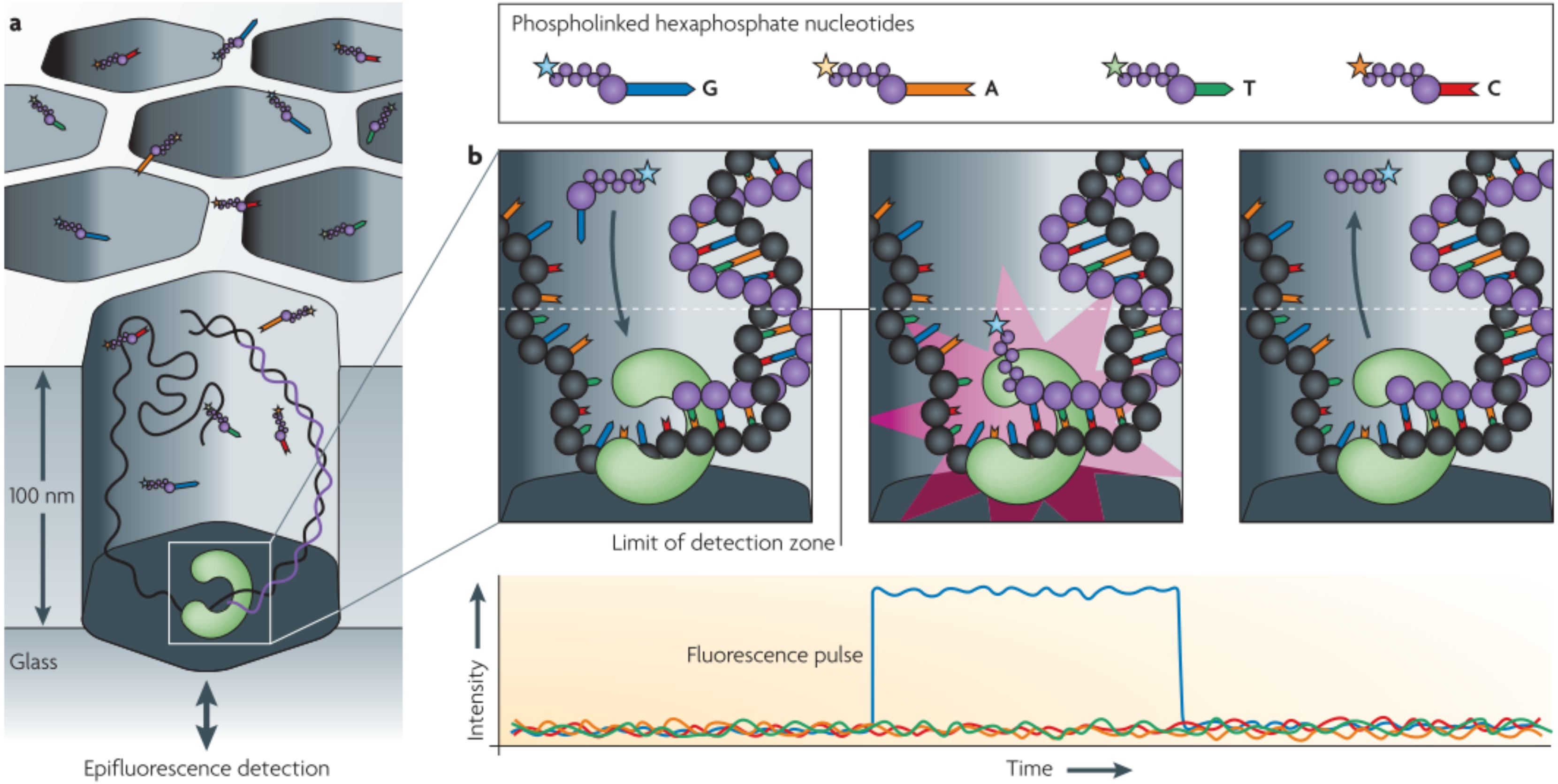
https://flxlexblog.wordpress.com/tag/illumina/

Third Generation Sequencing

- Third generation sequencing works by reading the nucleotide sequences at the single molecule level.
- Existing methods require breaking long strands of DNA into small segments then inferring nucleotide sequences by amplification and synthesis.
- Third Generation Sequencing = long-read sequencing
- PacBio and Oxford Nanopore provide manufactures

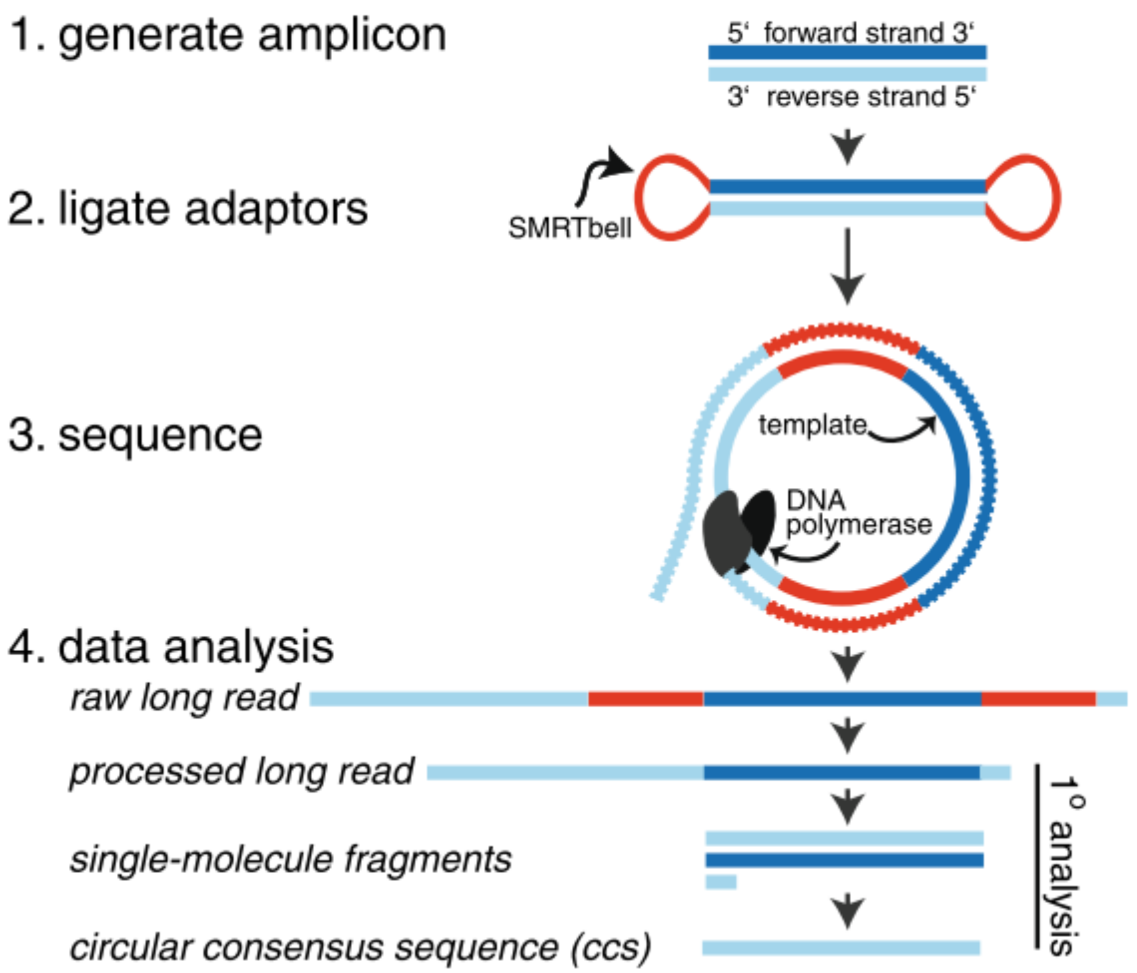
***Nature biotechnology*, 2008**

PacBio



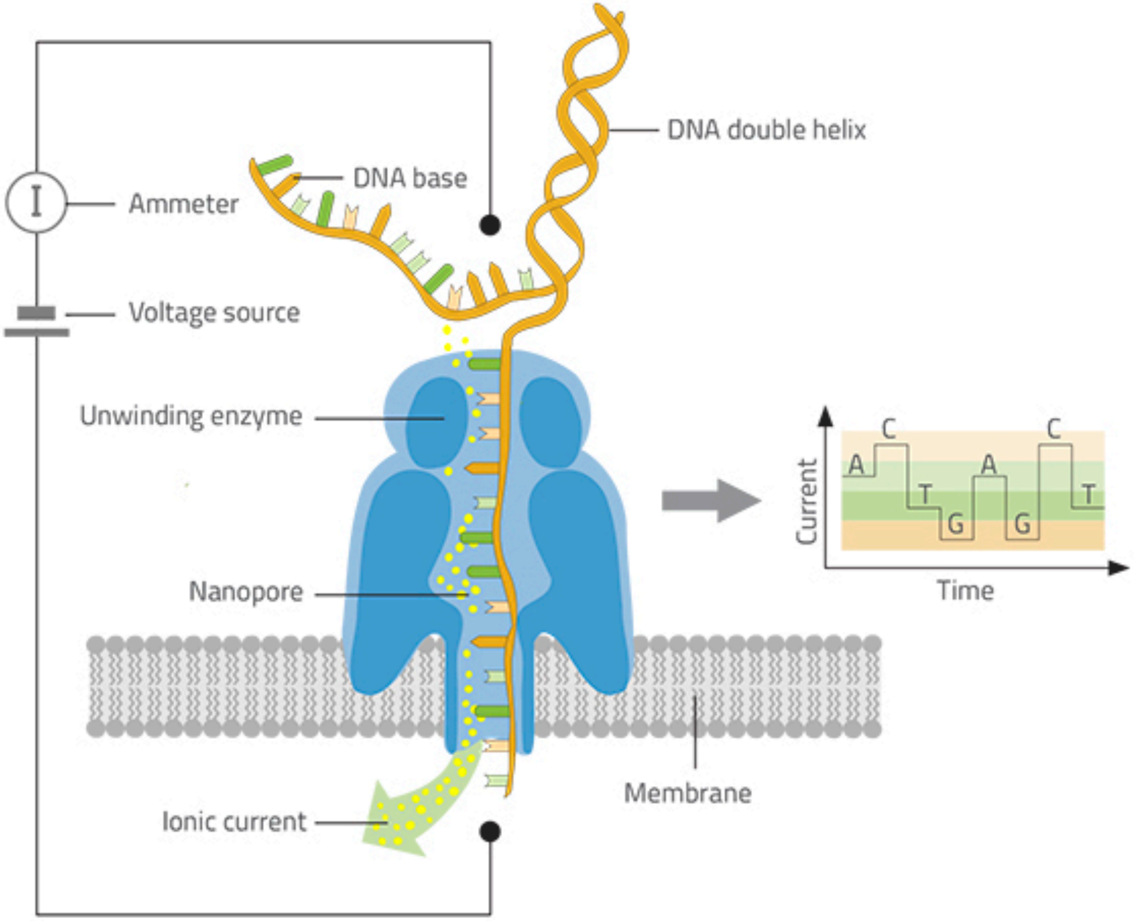
Nature reviews. Genetics, 2011

Steps in PacBio



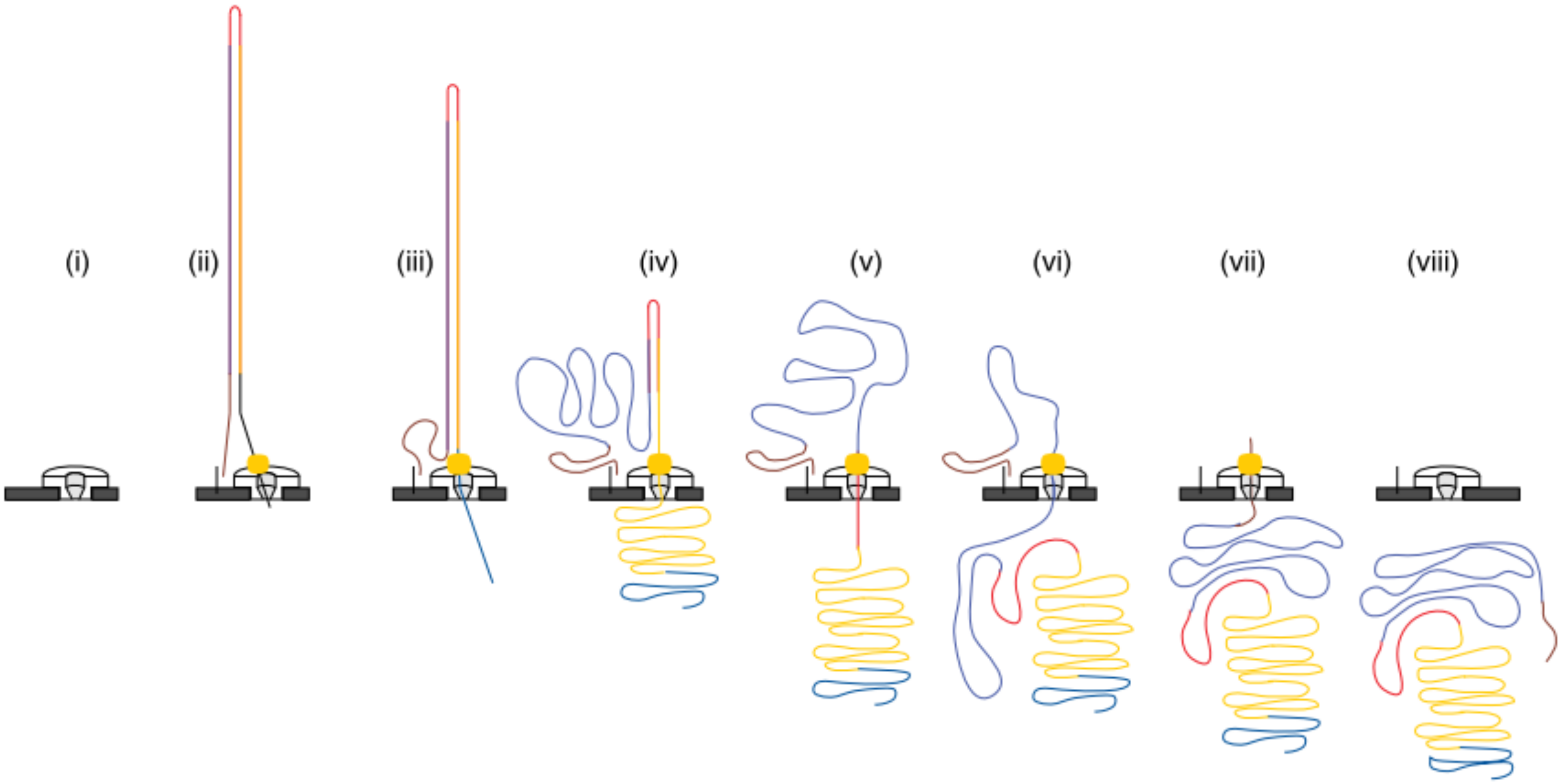
***Microbiome*, 2013**

Nanopore



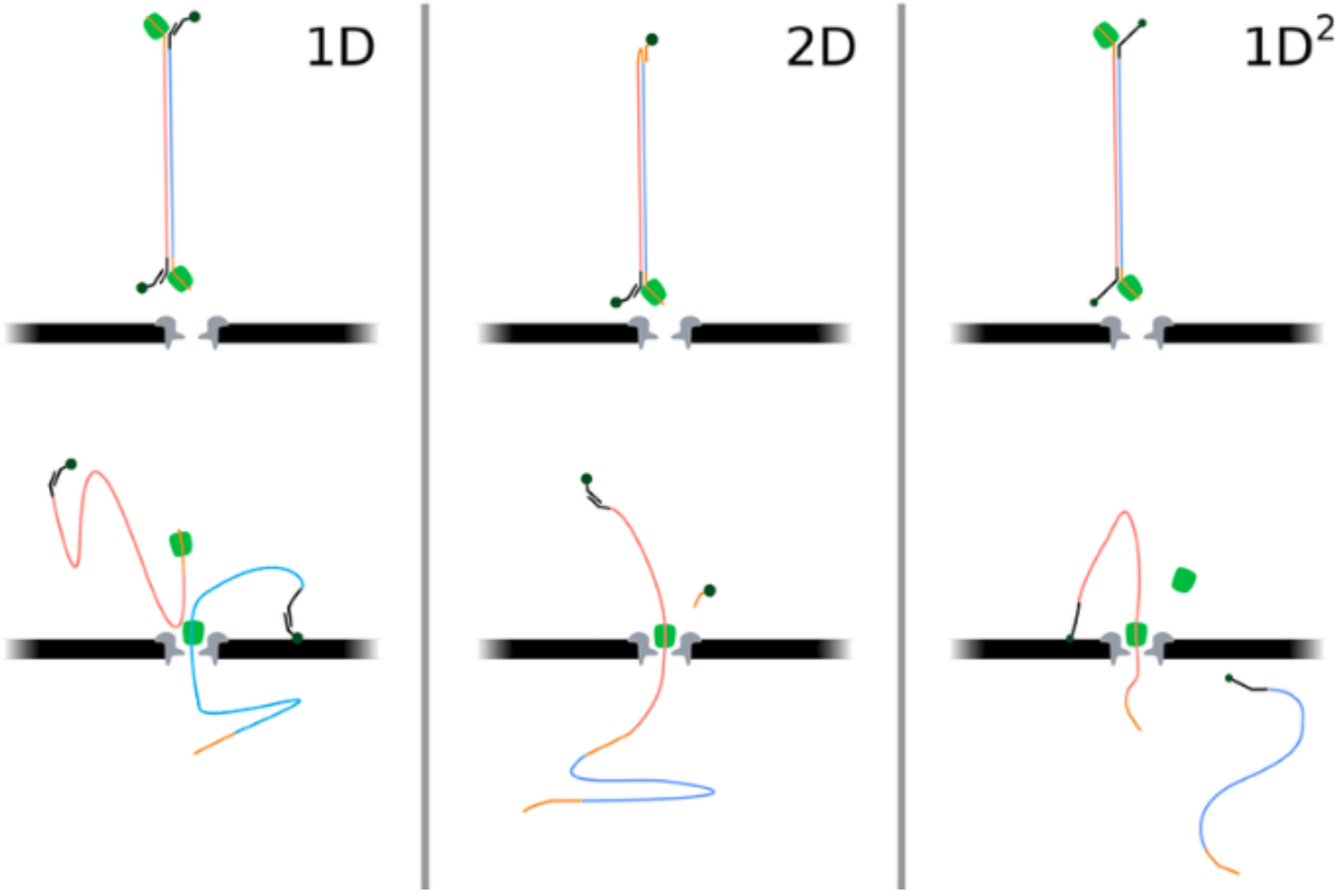
https://www.scienceinschool.org/content/decoding-dna-pocket-sized-sequencer

Steps in Nanopre



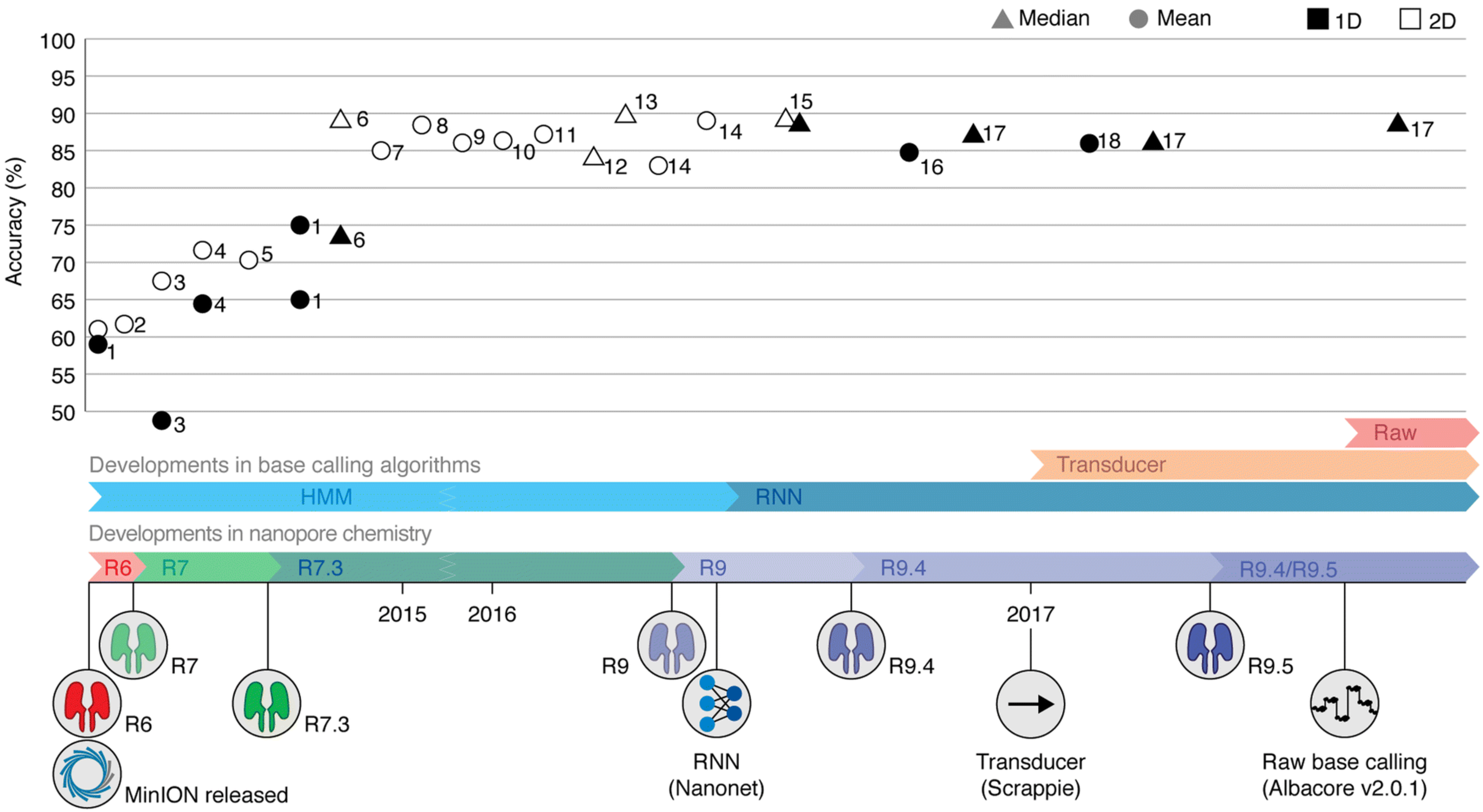
***Genome Biology*, 2011**

Strategy to improve sequencing accuracy



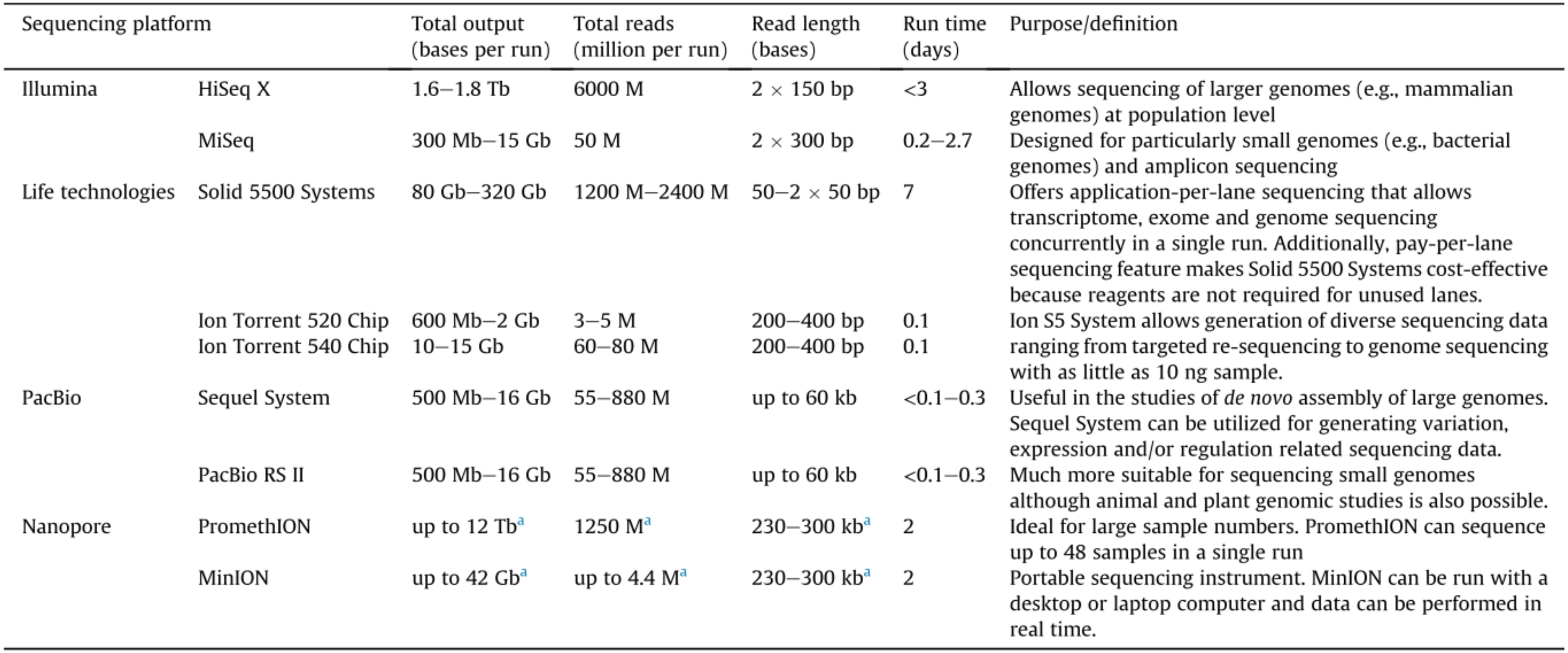
***Genome Biology*, 2011**

Accuracy of Nanopore



***Nature biotechnology*, 2008**

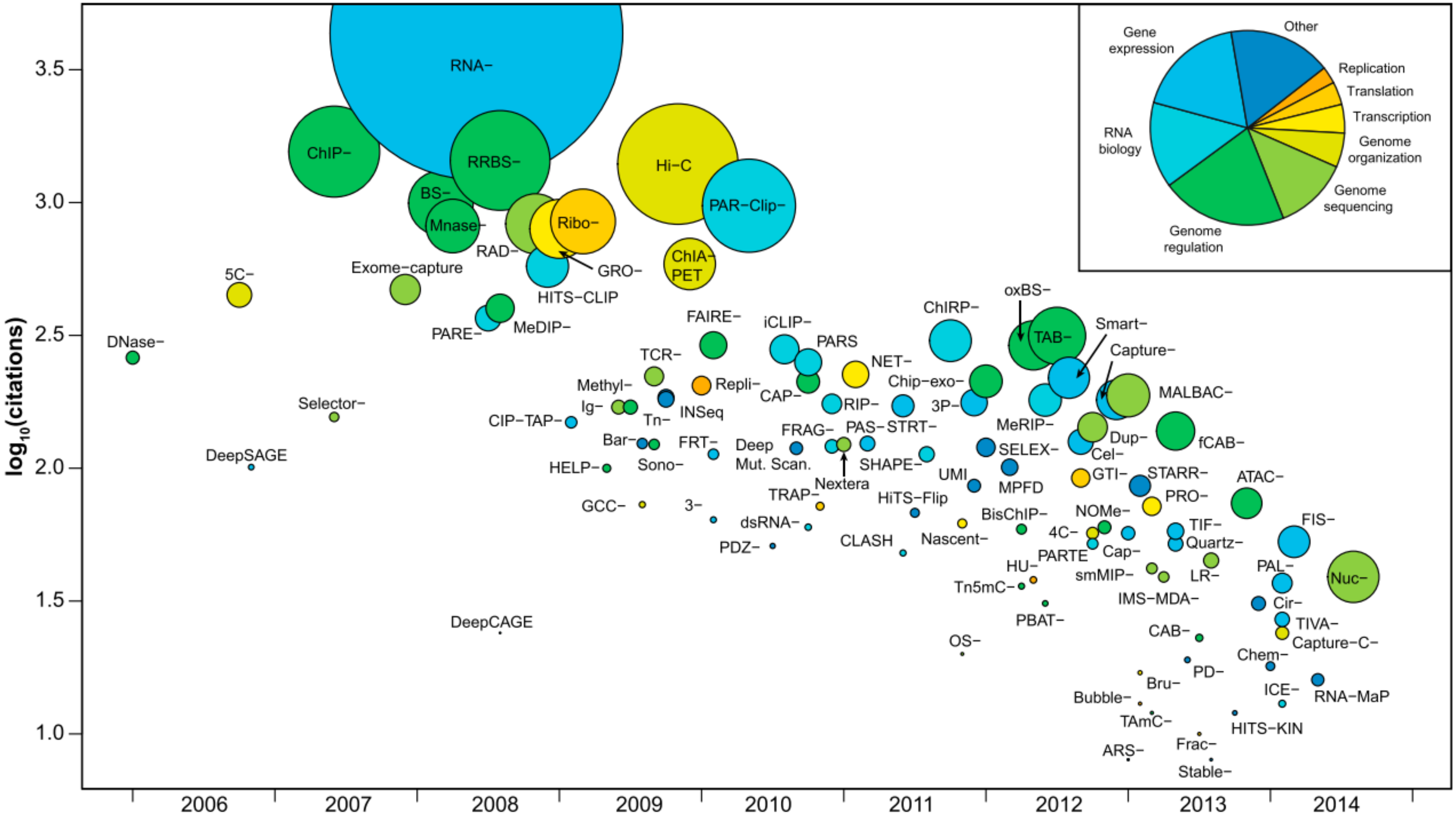
Comparison of platform between 2nd and 3rd



Sequencing platform		Total output (bases per run)	Total reads (million per run)	Read length (bases)	Run time (days)	Purpose/definition
Illumina	HiSeq X	1.6–1.8 Tb	6000 M	2 × 150 bp	<3	Allows sequencing of larger genomes (e.g., mammalian genomes) at population level
	MiSeq	300 Mb–15 Gb	50 M	2 × 300 bp	0.2–2.7	Designed for particularly small genomes (e.g., bacterial genomes) and amplicon sequencing
Life technologies	Solid 5500 Systems	80 Gb–320 Gb	1200 M–2400 M	50–2 × 50 bp	7	Offers application-per-lane sequencing that allows transcriptome, exome and genome sequencing concurrently in a single run. Additionally, pay-per-lane sequencing feature makes Solid 5500 Systems cost-effective because reagents are not required for unused lanes.
	Ion Torrent 520 Chip	600 Mb–2 Gb	3–5 M	200–400 bp	0.1	Ion S5 System allows generation of diverse sequencing data ranging from targeted re-sequencing to genome sequencing with as little as 10 ng sample.
	Ion Torrent 540 Chip	10–15 Gb	60–80 M	200–400 bp	0.1	
PacBio	Sequel System	500 Mb–16 Gb	55–880 M	up to 60 kb	<0.1–0.3	Useful in the studies of *de novo* assembly of large genomes. Sequel System can be utilized for generating variation, expression and/or regulation related sequencing data.
	PacBio RS II	500 Mb–16 Gb	55–880 M	up to 60 kb	<0.1–0.3	Much more suitable for sequencing small genomes although animal and plant genomic studies is also possible.
Nanopore	PromethION	up to 12 Tb[a]	1250 M[a]	230–300 kb[a]	2	Ideal for large sample numbers. PromethION can sequence up to 48 samples in a single run
	MinION	up to 42 Gb[a]	up to 4.4 M[a]	230–300 kb[a]	2	Portable sequencing instrument. MinION can be run with a desktop or laptop computer and data can be performed in real time.

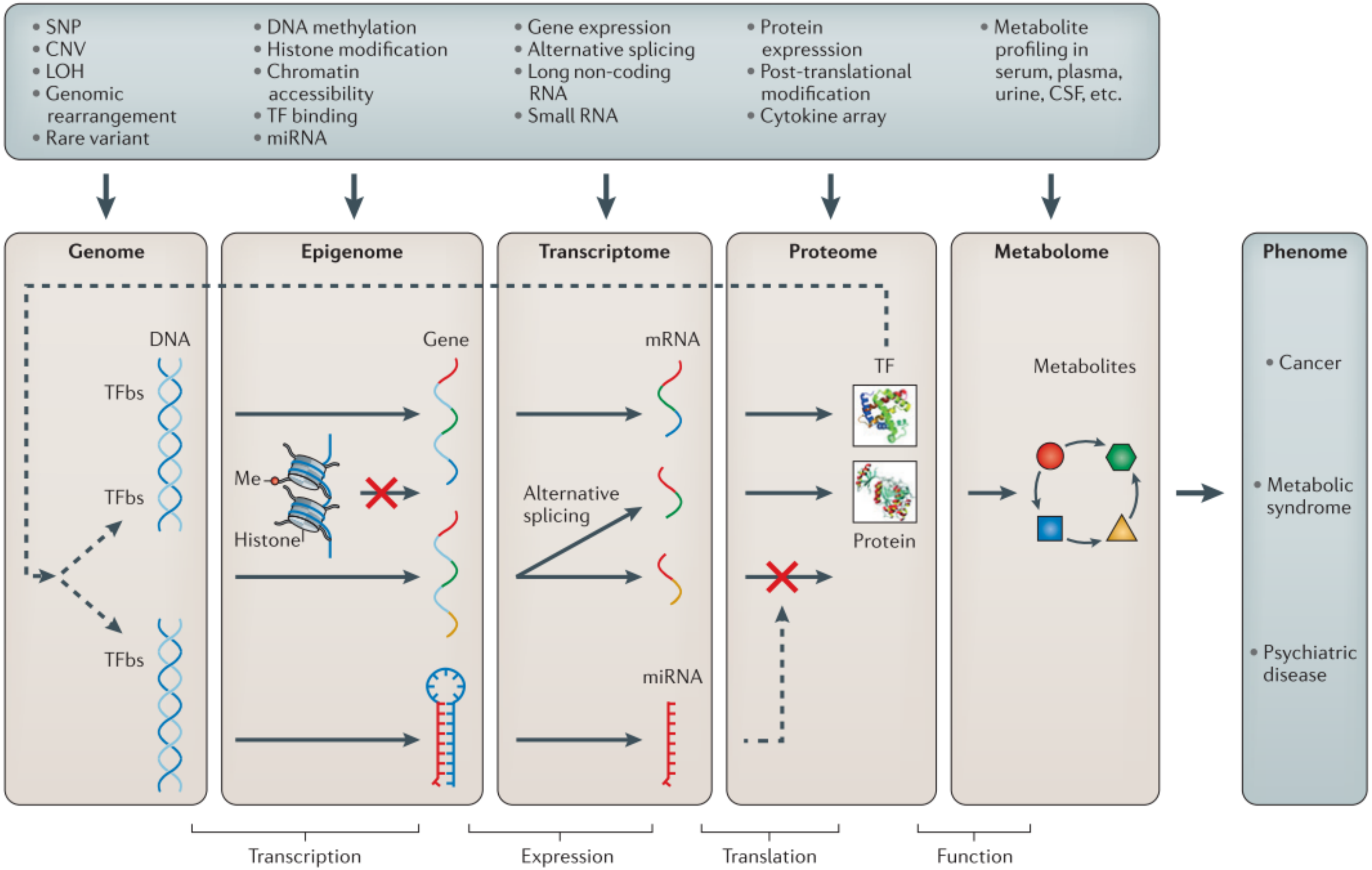
***Progress in Retinal and Eye Research*, 2016**

Application of NGS



Molecular cell, 2015

Biological layer



- SNP
- CNV
- LOH
- Genomic rearrangement
- Rare variant

- DNA methylation
- Histone modification
- Chromatin accessibility
- TF binding
- miRNA

- Gene expression
- Alternative splicing
- Long non-coding RNA
- Small RNA

- Protein expresssion
- Post-translational modification
- Cytokine array

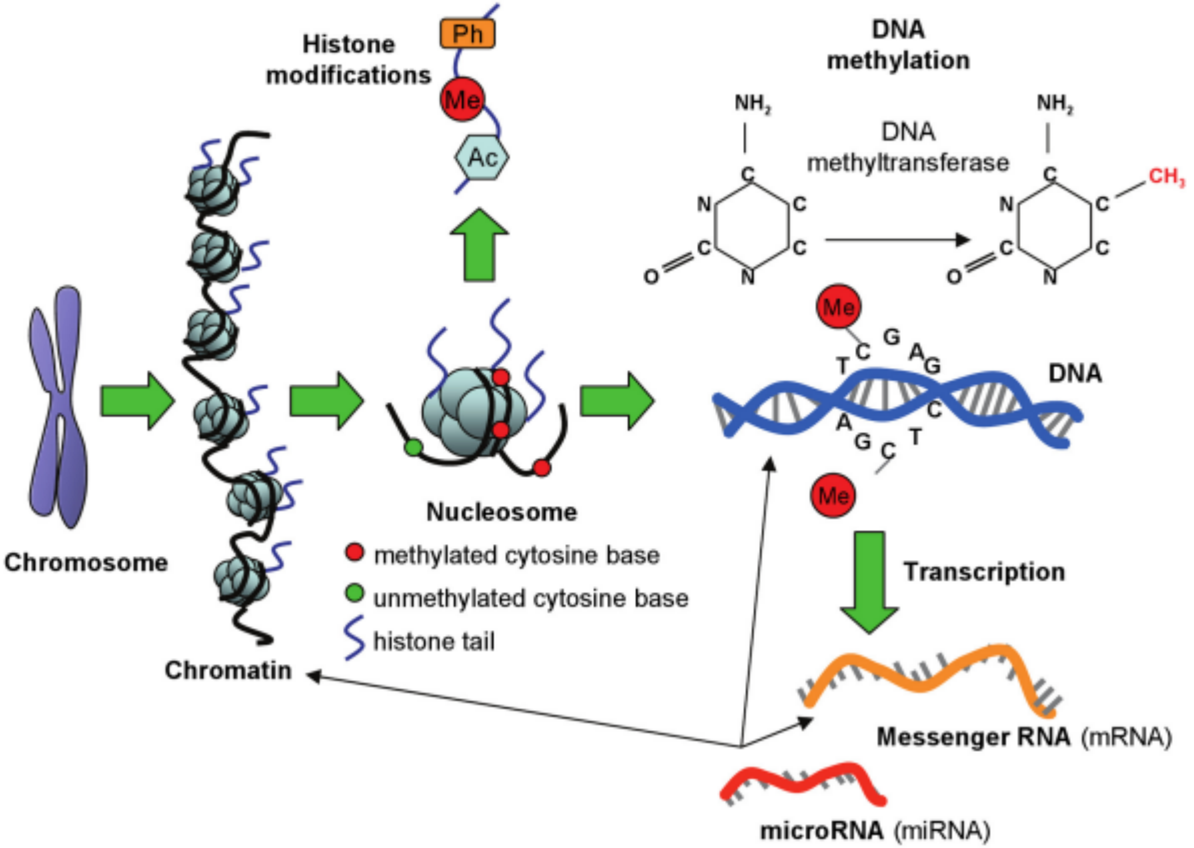
- Metabolite profiling in serum, plasma, urine, CSF, etc.

Genome → **Epigenome** → **Transcriptome** → **Proteome** → **Metabolome** → **Phenome**

- Cancer
- Metabolic syndrome
- Psychiatric disease

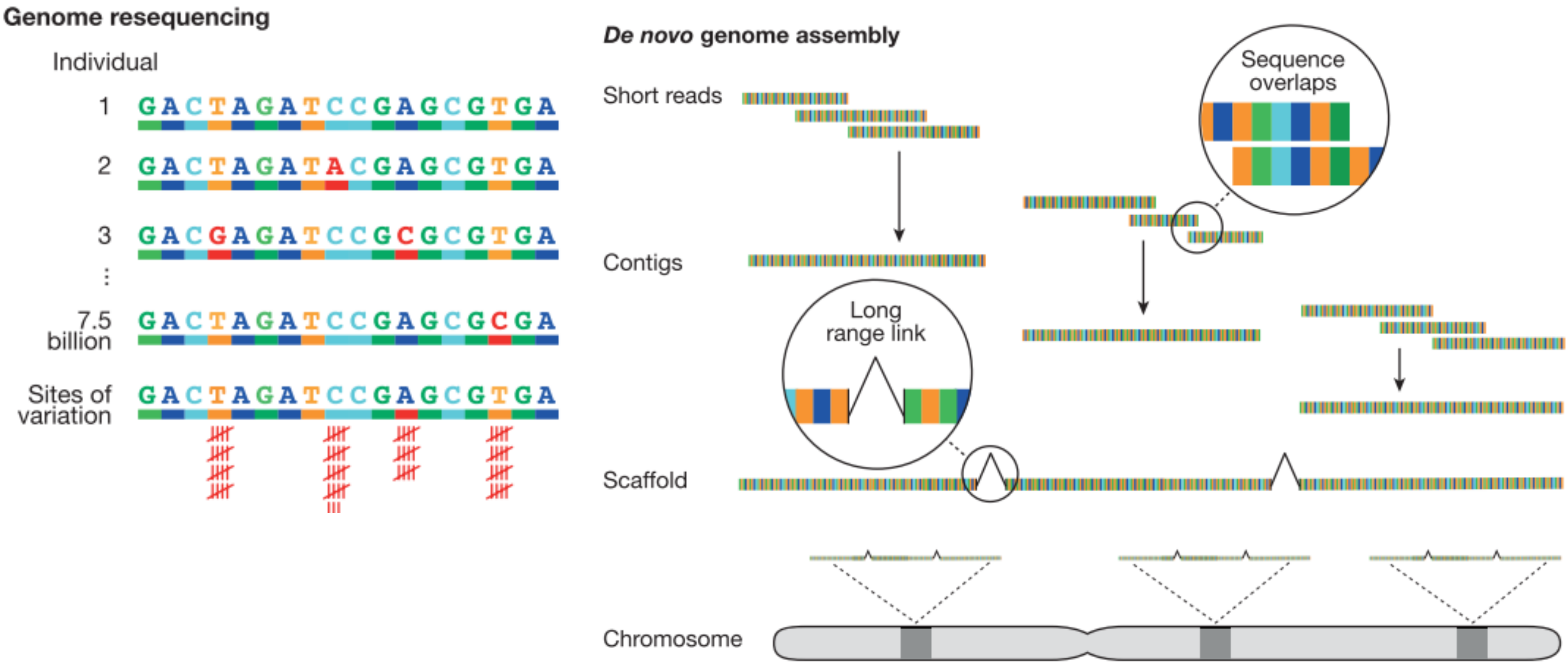
***Nature Reviews Genetics*, 2015**

Central dogma with epigenome



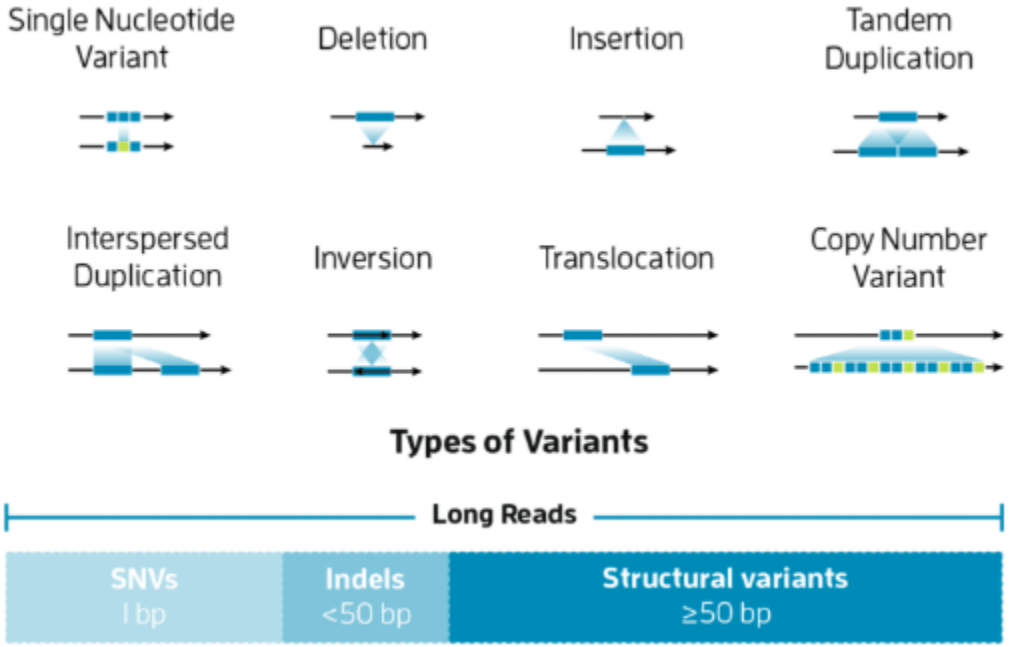
***PLoS Medicine*, 2010**

Sequencing using whole genome



***Nature*, 2017**

Type of variants



https://www.pacb.com/applications/whole-genome-sequencing/variant-detection/

The Angelina effect

MAY 27, 2013

The IRS Mess / Syria's YouTube War / The End of Alimony

TIME

THE ANGELINA EFFECT

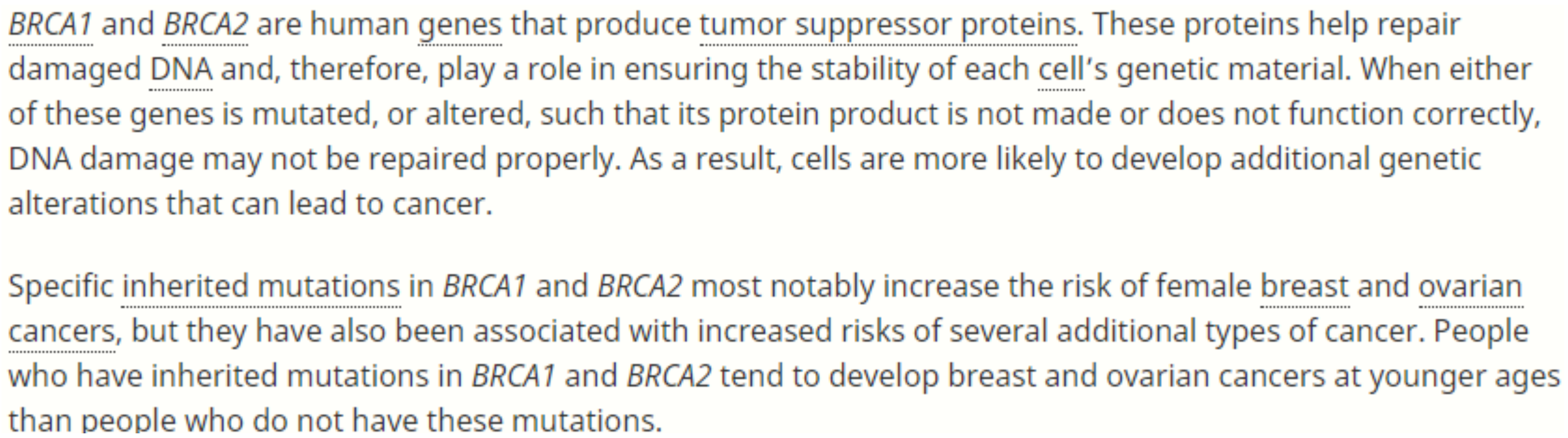
Angelina Jolie's double mastectomy puts genetic testing in the spotlight. What her choice reveals about calculating risk, cost and peace of mind

BY JEFFREY KLUGER & ALICE PARK

time.com

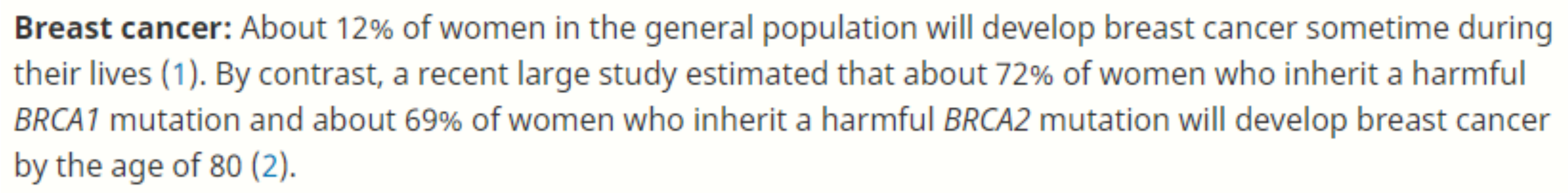
Time Magazine

BRCA1 and BRCA2

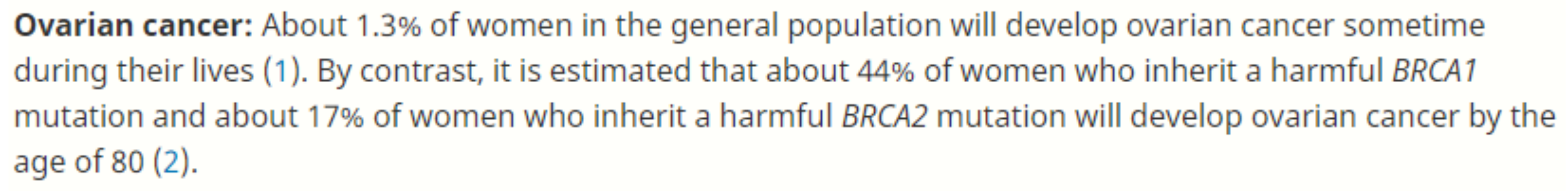


BRCA1 and *BRCA2* are human genes that produce tumor suppressor proteins. These proteins help repair damaged DNA and, therefore, play a role in ensuring the stability of each cell's genetic material. When either of these genes is mutated, or altered, such that its protein product is not made or does not function correctly, DNA damage may not be repaired properly. As a result, cells are more likely to develop additional genetic alterations that can lead to cancer.

Specific inherited mutations in *BRCA1* and *BRCA2* most notably increase the risk of female breast and ovarian cancers, but they have also been associated with increased risks of several additional types of cancer. People who have inherited mutations in *BRCA1* and *BRCA2* tend to develop breast and ovarian cancers at younger ages than people who do not have these mutations.



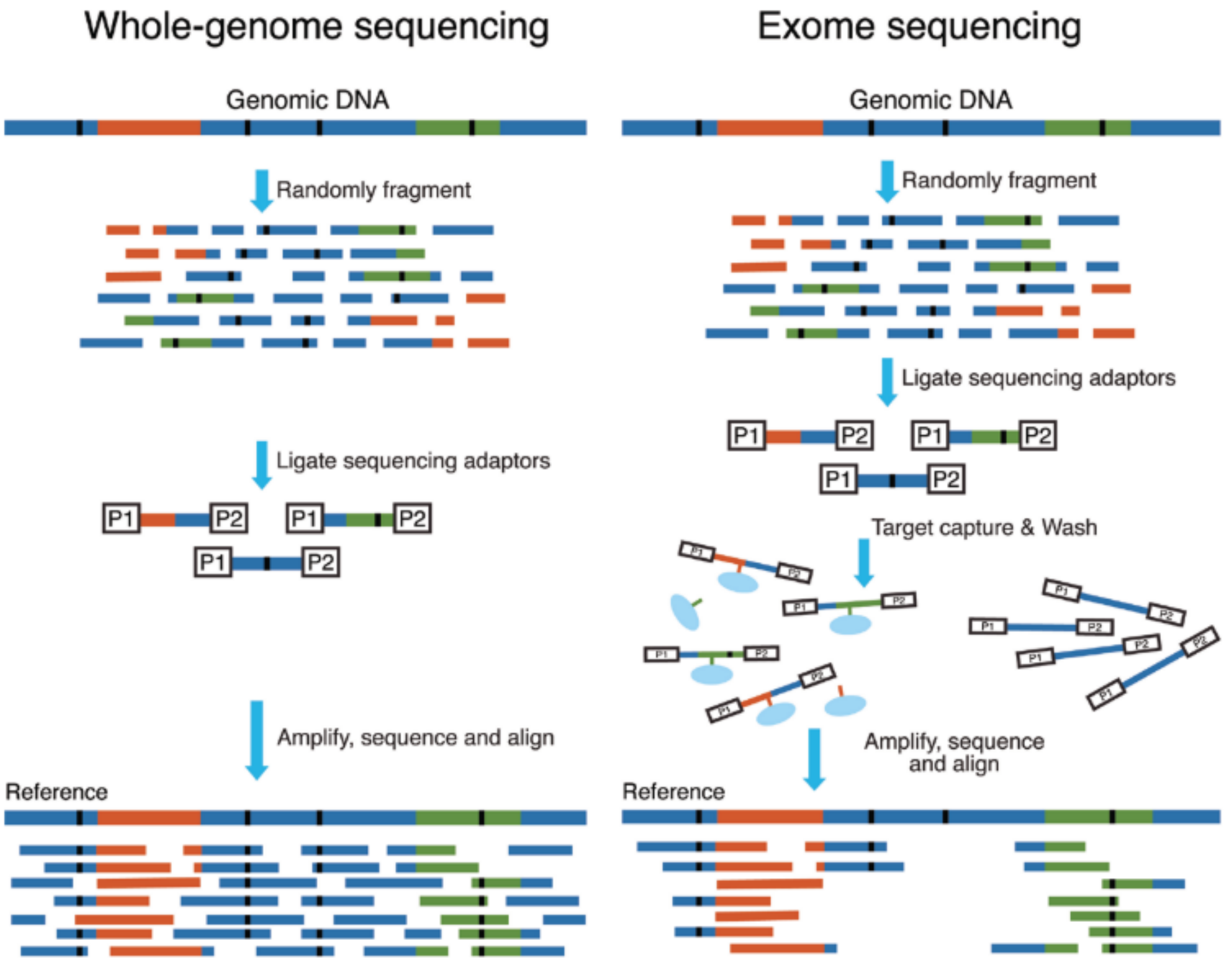
Breast cancer: About 12% of women in the general population will develop breast cancer sometime during their lives (1). By contrast, a recent large study estimated that about 72% of women who inherit a harmful *BRCA1* mutation and about 69% of women who inherit a harmful *BRCA2* mutation will develop breast cancer by the age of 80 (2).



Ovarian cancer: About 1.3% of women in the general population will develop ovarian cancer sometime during their lives (1). By contrast, it is estimated that about 44% of women who inherit a harmful *BRCA1* mutation and about 17% of women who inherit a harmful *BRCA2* mutation will develop ovarian cancer by the age of 80 (2).

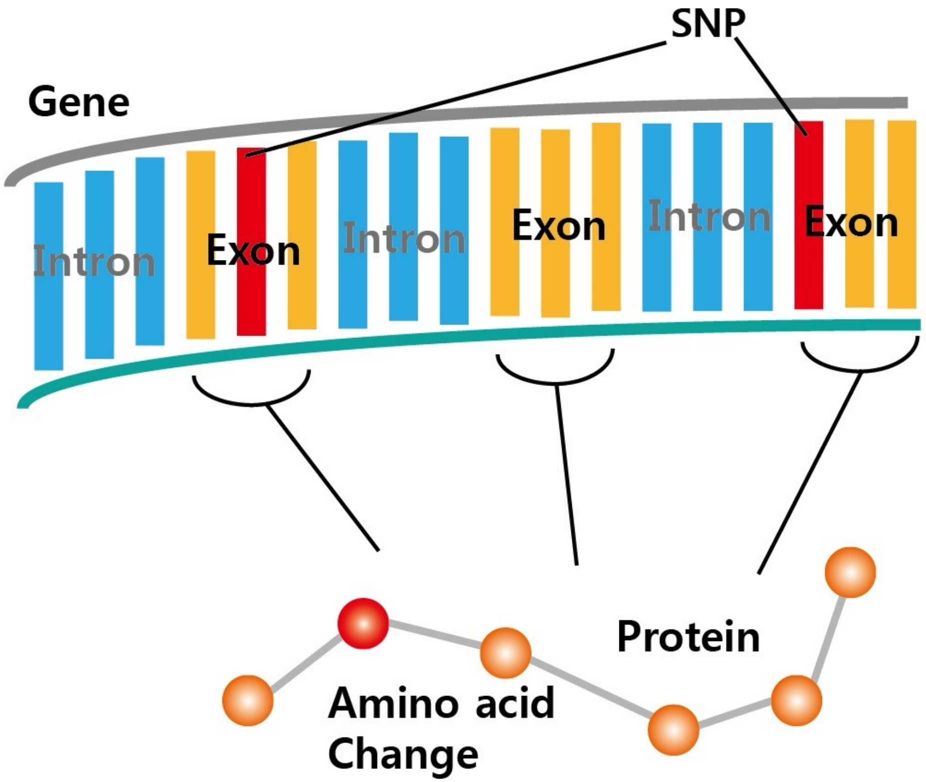
https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet

Sequencing for resequencing



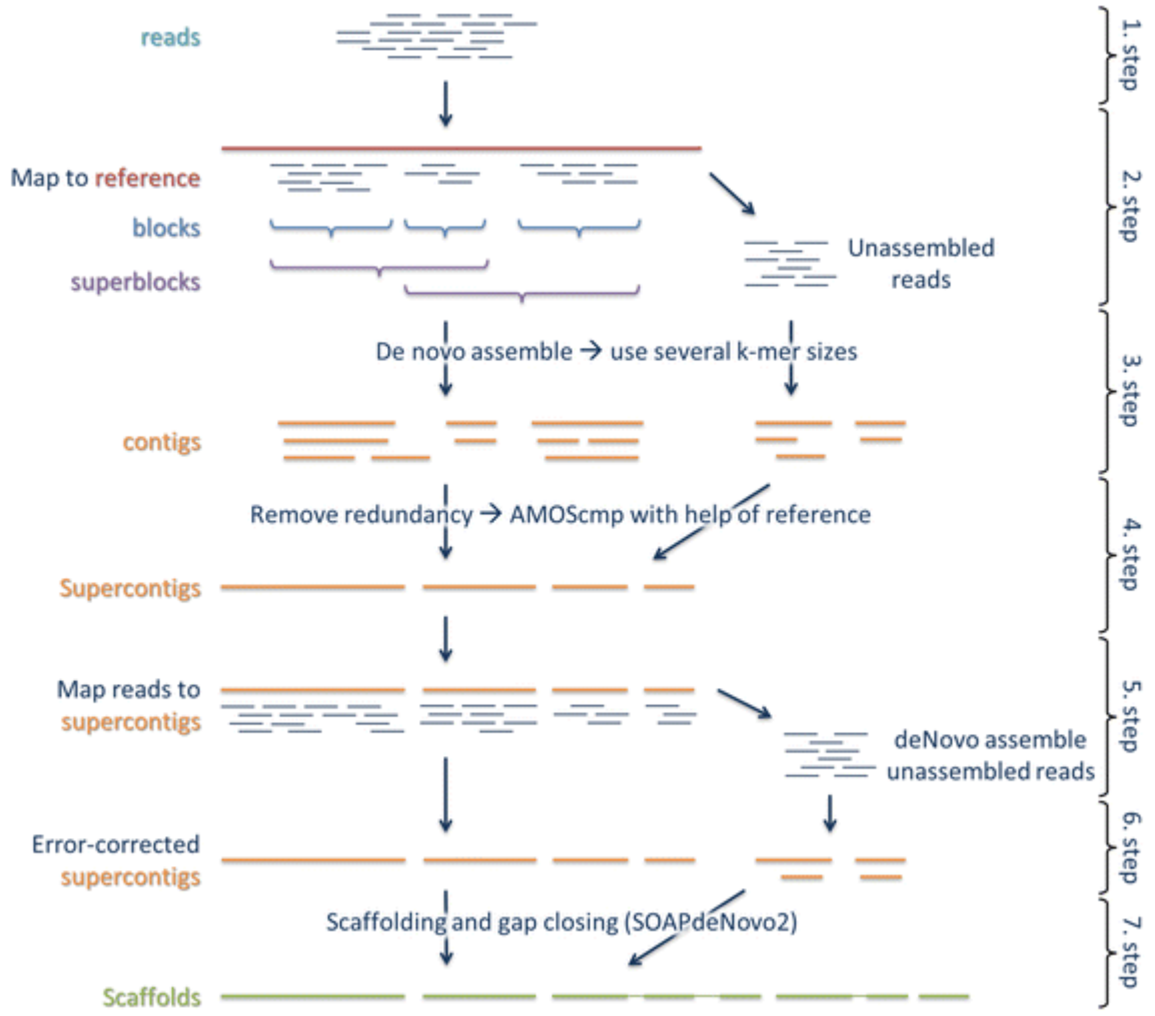
https://ngxbio.com/pub/media/magefan_blog/WES1.jpg

Whole Exome Sequencing



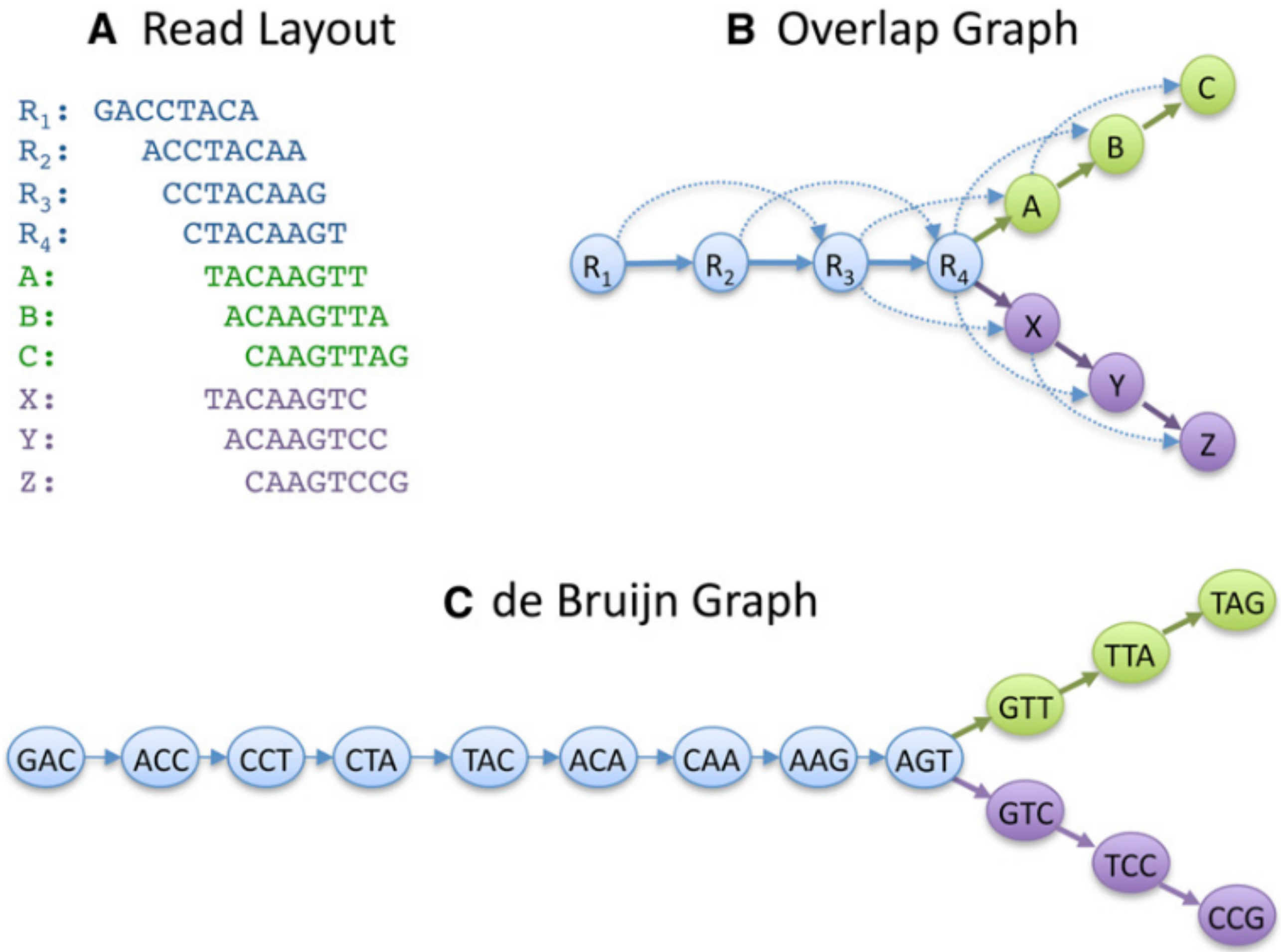
https://ngxbio.com/pub/media/magefan_blog/WES1.jpg

Steps of De novo genome assembly


reads
Map to reference
blocks
superblocks
Unassembled reads
De novo assemble → use several k-mer sizes
contigs
Remove redundancy → AMOScmp with help of reference
Supercontigs
Map reads to supercontigs
deNovo assemble unassembled reads
Error-corrected supercontigs
Scaffolding and gap closing (SOAPdeNovo2)
Scaffolds
1. step
2. step
3. step
4. step
5. step
6. step
7. step

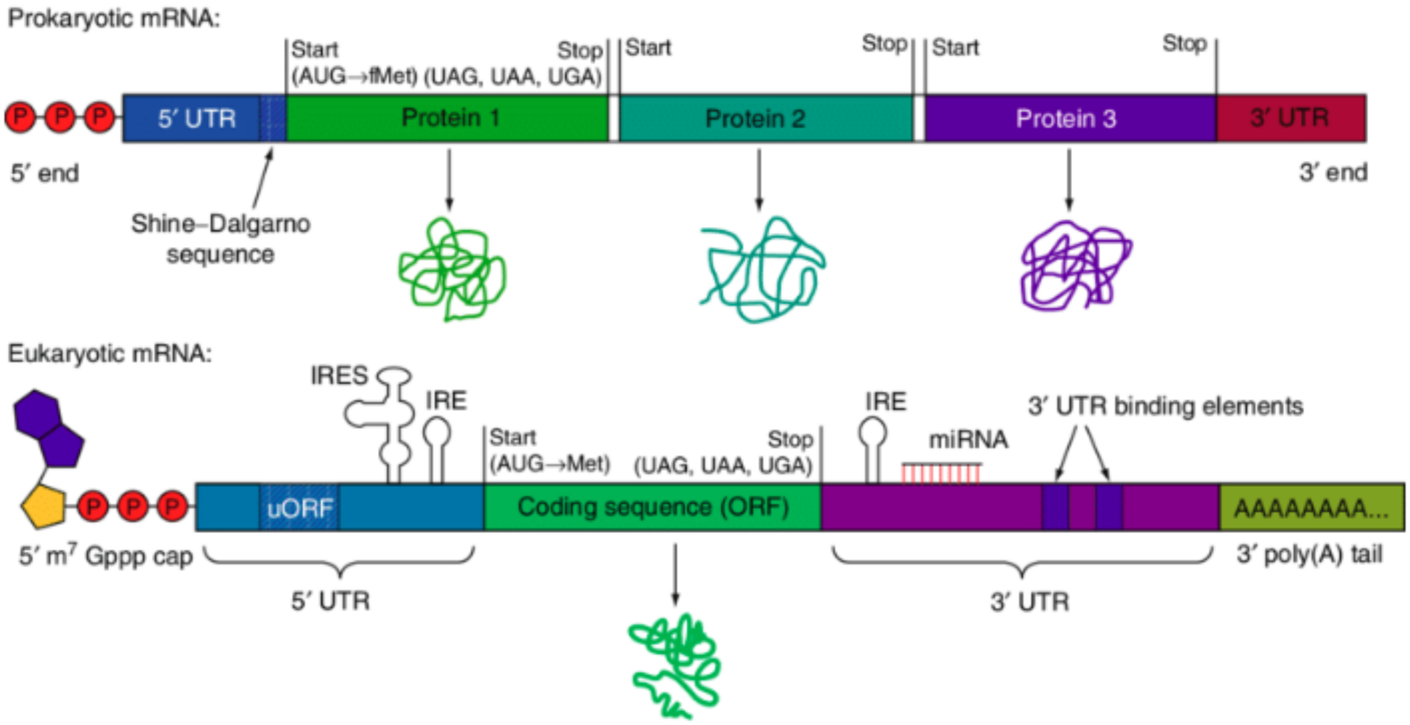
***BMC Bioinformatics*, 2017**

Genome assembly algorithms



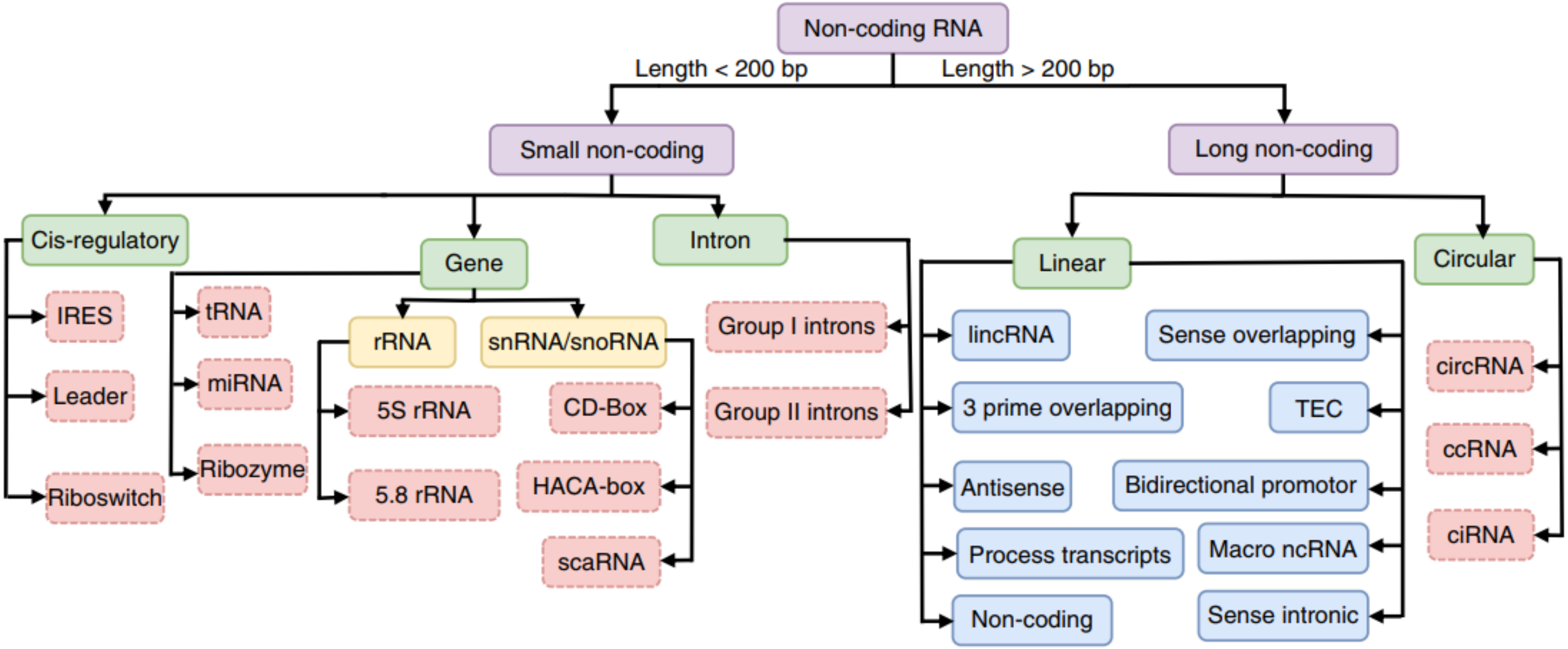
***Genome research*, 2010**

Messenger RNA



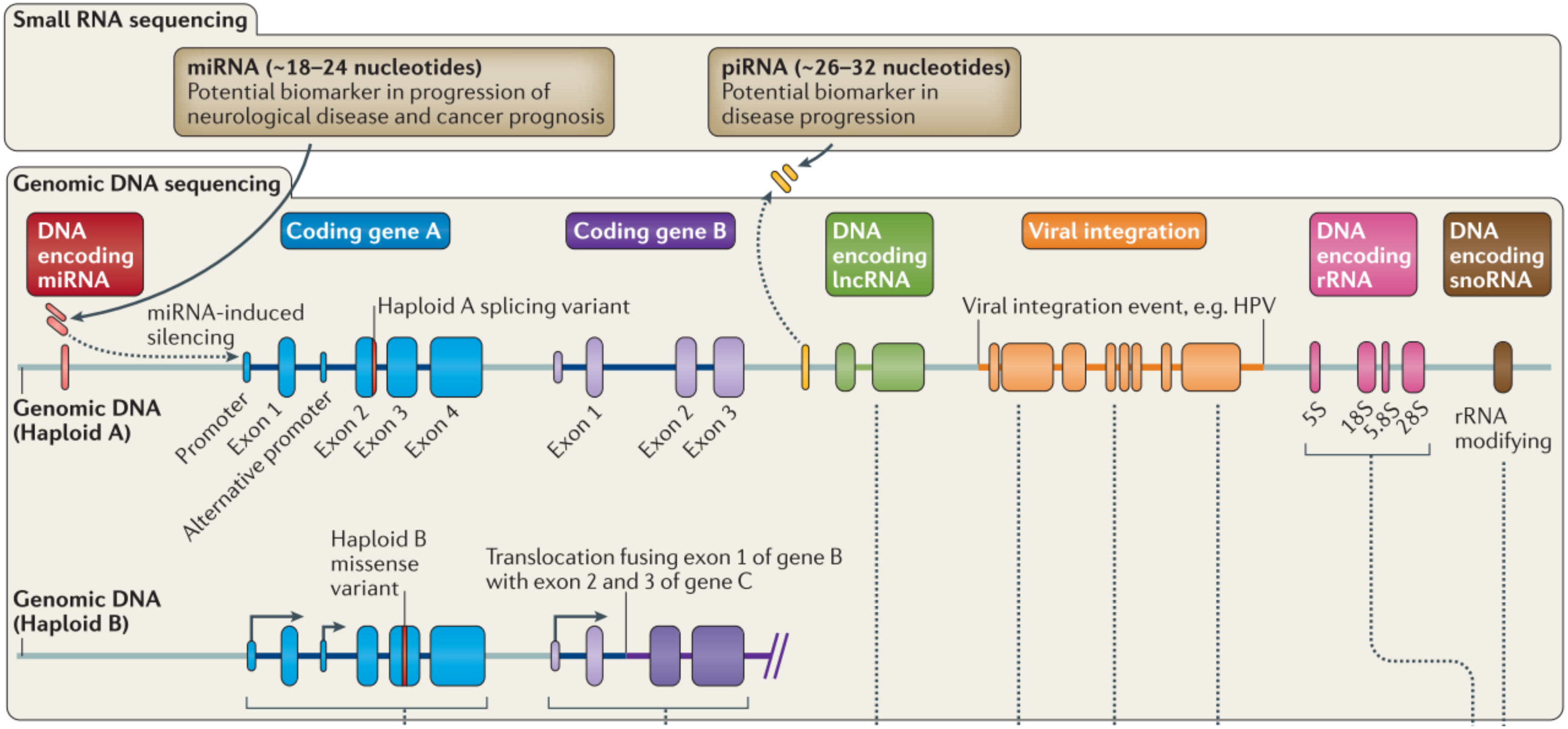
https://www.researchgate.net/figure/Schematic-diagram-of-prokaryotic-top-and-eukaryotic-bottom-mRNA-Bars-indicate-the_fig1_305698915

Non-coding RNA



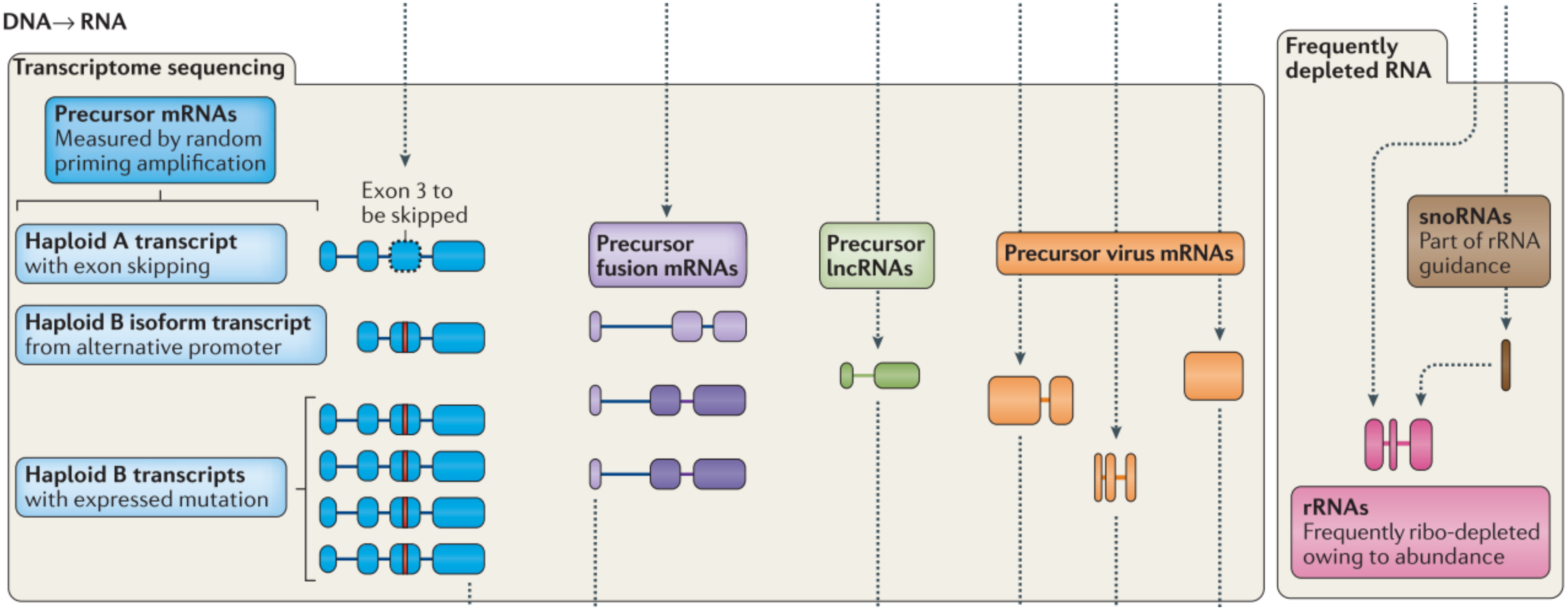
***Genome research*, 2010**

Type of RNA-seq


Small RNA sequencing
miRNA (~18–24 nucleotides)
Potential biomarker in progression of neurological disease and cancer prognosis
piRNA (~26–32 nucleotides)
Potential biomarker in disease progression
Genomic DNA sequencing
DNA encoding miRNA
Coding gene A
Coding gene B
DNA encoding lncRNA
Viral integration
DNA encoding rRNA
DNA encoding snoRNA
miRNA-induced silencing
Haploid A splicing variant
Viral integration event, e.g. HPV
Genomic DNA (Haploid A)
Promoter
Exon 1
Alternative promoter
Exon 2
Exon 3
Exon 4
Exon 1
Exon 2
Exon 3
5S
18S
5.8S
28S
rRNA modifying
Haploid B missense variant
Translocation fusing exon 1 of gene B with exon 2 and 3 of gene C
Genomic DNA (Haploid B)

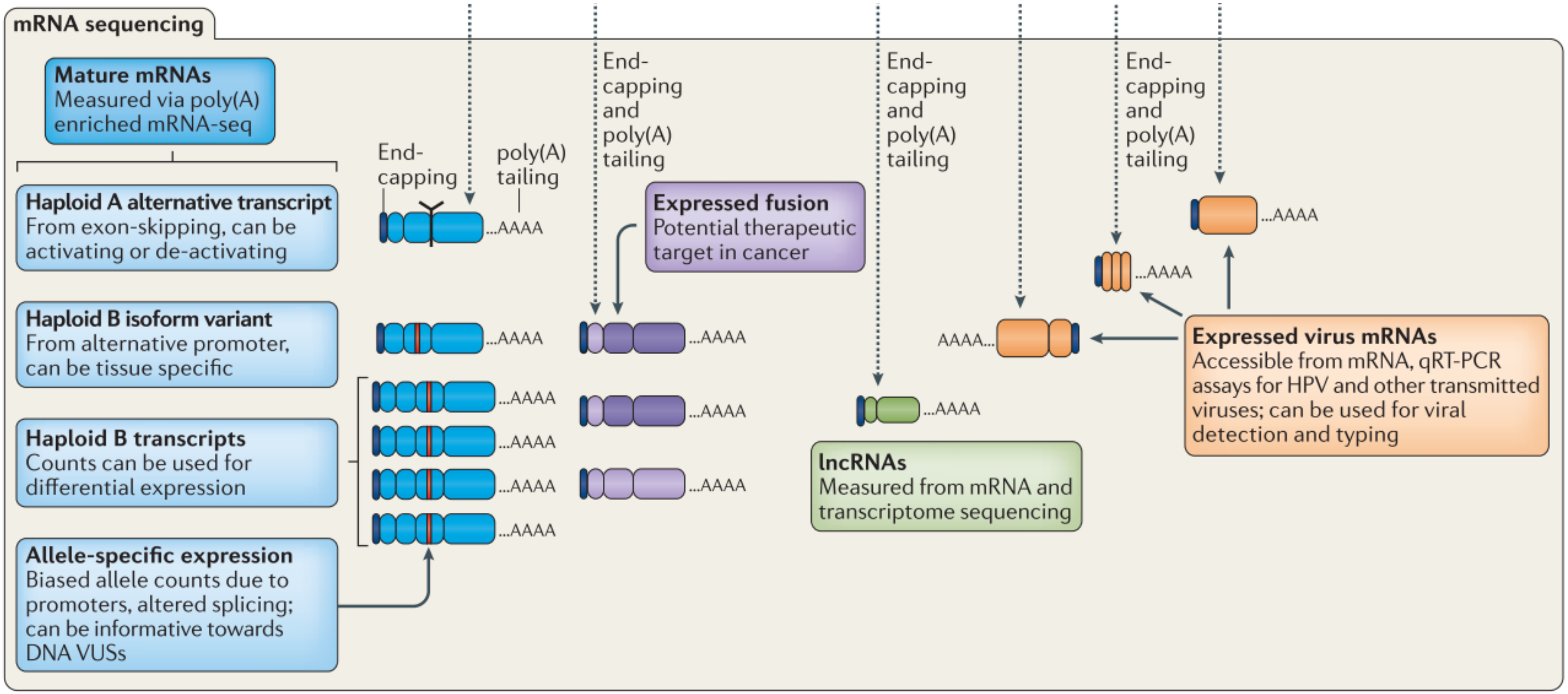
Nature Reviews Genetics, 2016

Type of RNA-seq



Nature Reviews Genetics, 2016

Type of RNA-seq



mRNA sequencing

Mature mRNAs
Measured via poly(A) enriched mRNA-seq

Haploid A alternative transcript
From exon-skipping, can be activating or de-activating

Haploid B isoform variant
From alternative promoter, can be tissue specific

Haploid B transcripts
Counts can be used for differential expression

Allele-specific expression
Biased allele counts due to promoters, altered splicing; can be informative towards DNA VUSs

Expressed fusion
Potential therapeutic target in cancer

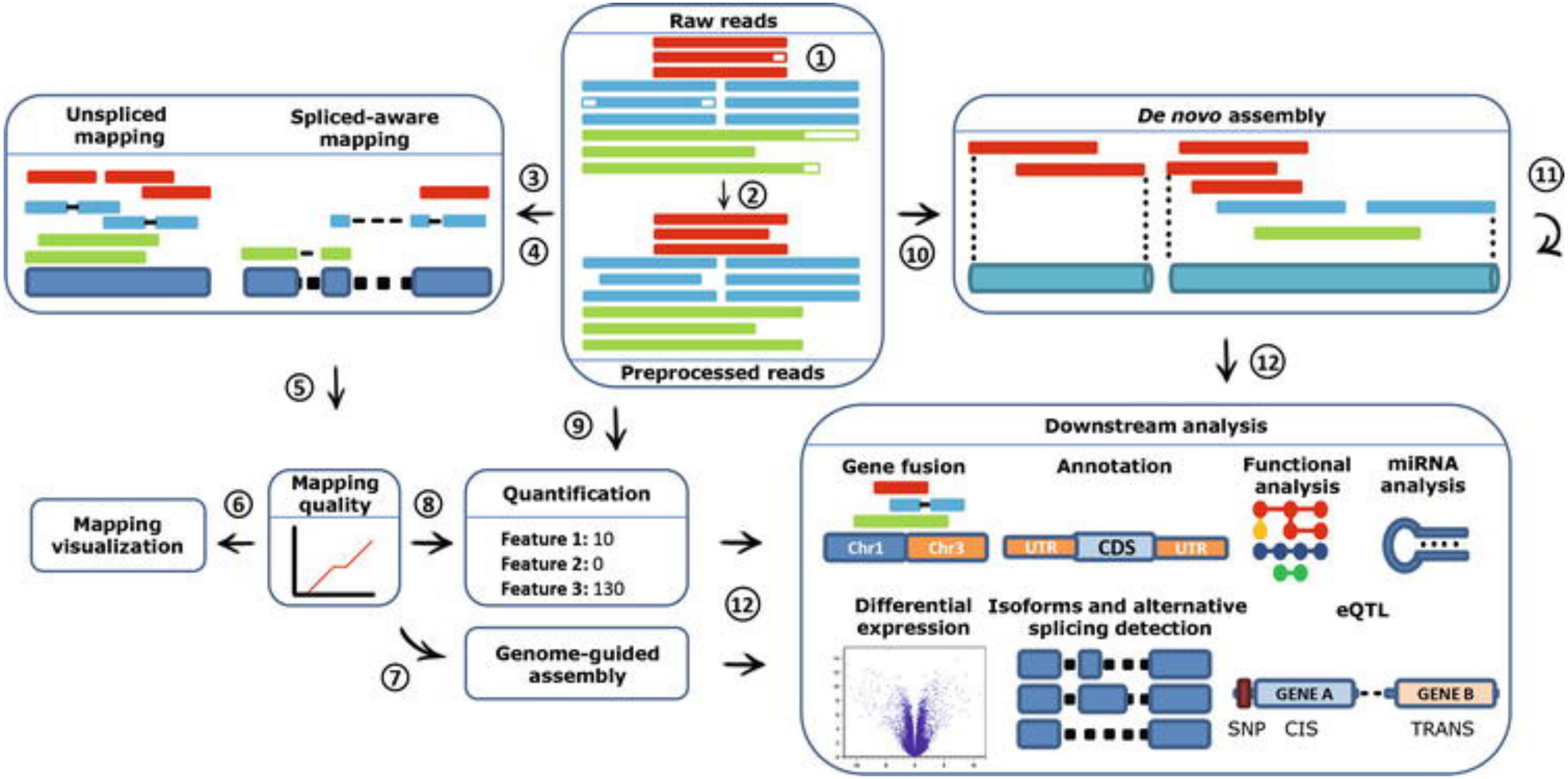
lncRNAs
Measured from mRNA and transcriptome sequencing

Expressed virus mRNAs
Accessible from mRNA, qRT-PCR assays for HPV and other transmitted viruses; can be used for viral detection and typing

End-capping
poly(A) tailing
End-capping and poly(A) tailing
End-capping and poly(A) tailing
End-capping and poly(A) tailing
...AAAA
AAAA...

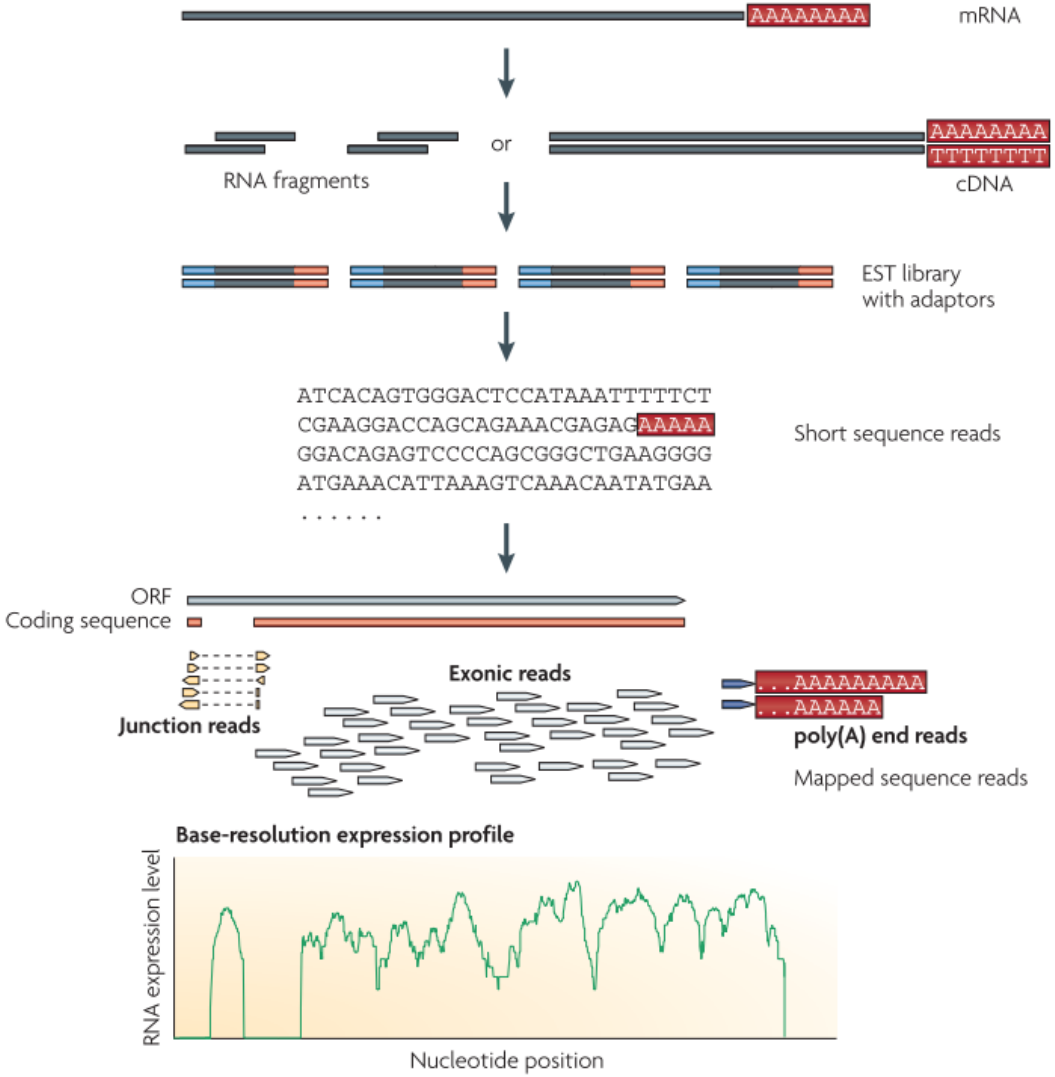
Nature Reviews Genetics, 2016

RNA-seq data analysis



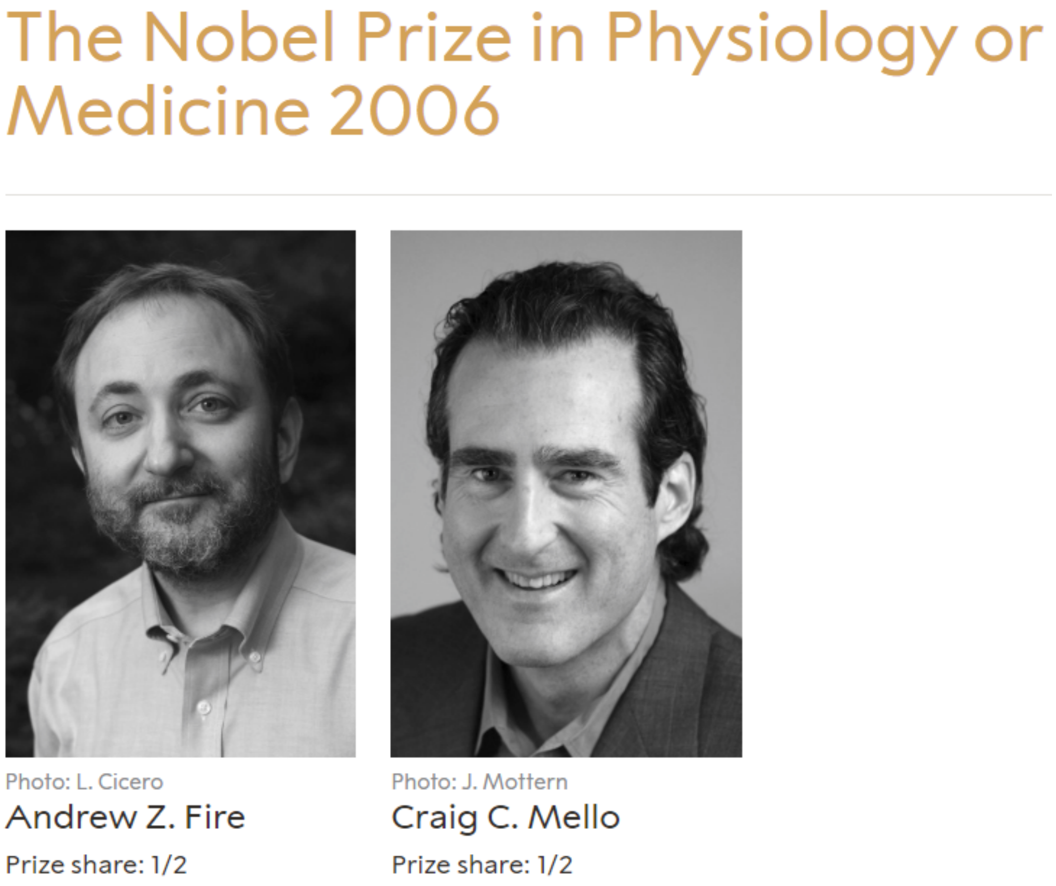
https://www.intechopen.com/books/applications-of-rna-seq-and-omics-strategies-from-microorganisms-to-human-health/rna-seq-applications-and-best-practices

RNA-seq data analysis



Nature reviews genetics, 2009

Nobel prize – RNA Interference



The Nobel Prize in Physiology or Medicine 2006

Photo: L. Cicero

Andrew Z. Fire

Prize share: 1/2

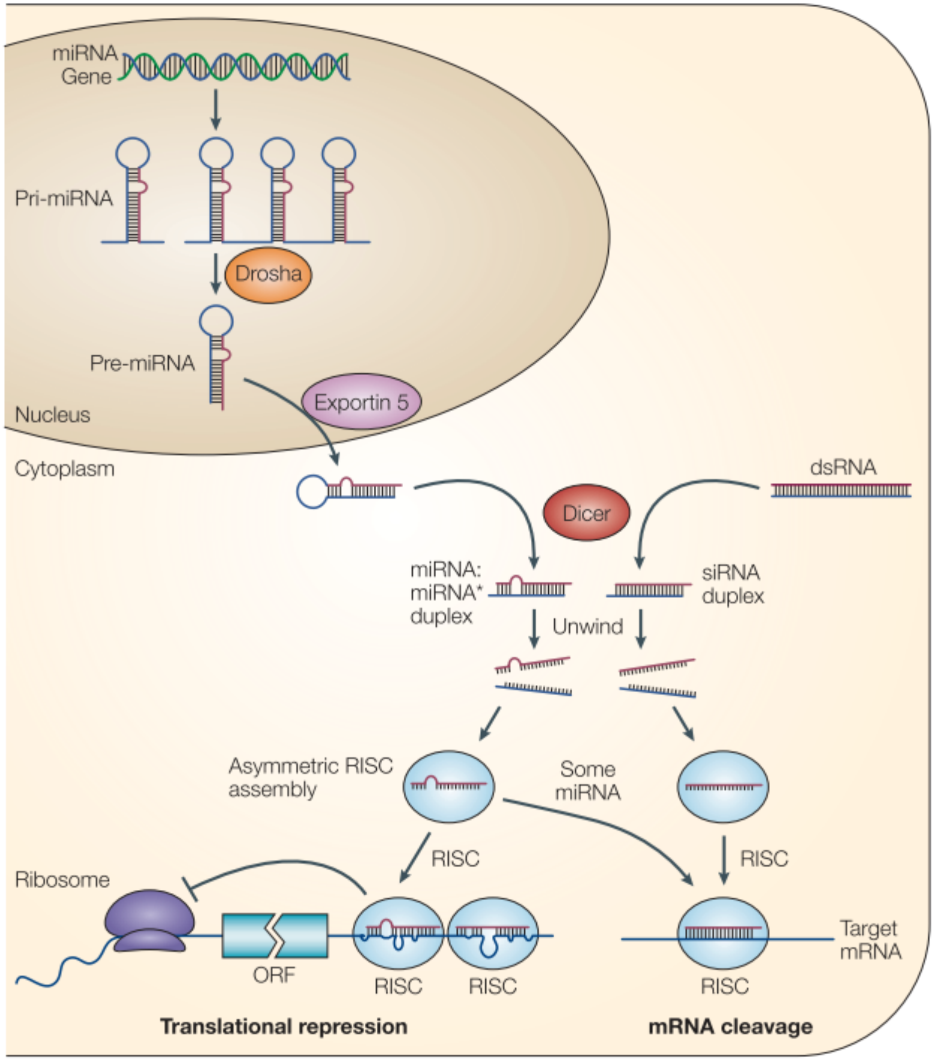
Photo: J. Mottern

Craig C. Mello

Prize share: 1/2

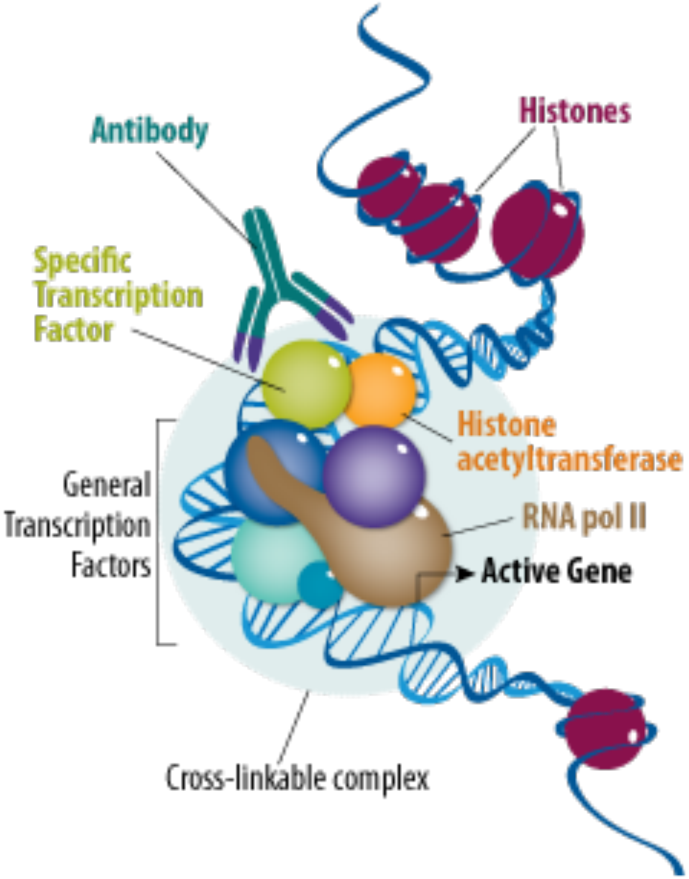
https://www.nobelprize.org/prizes/medicine/2006/summary/

Micro RNA biogenesis



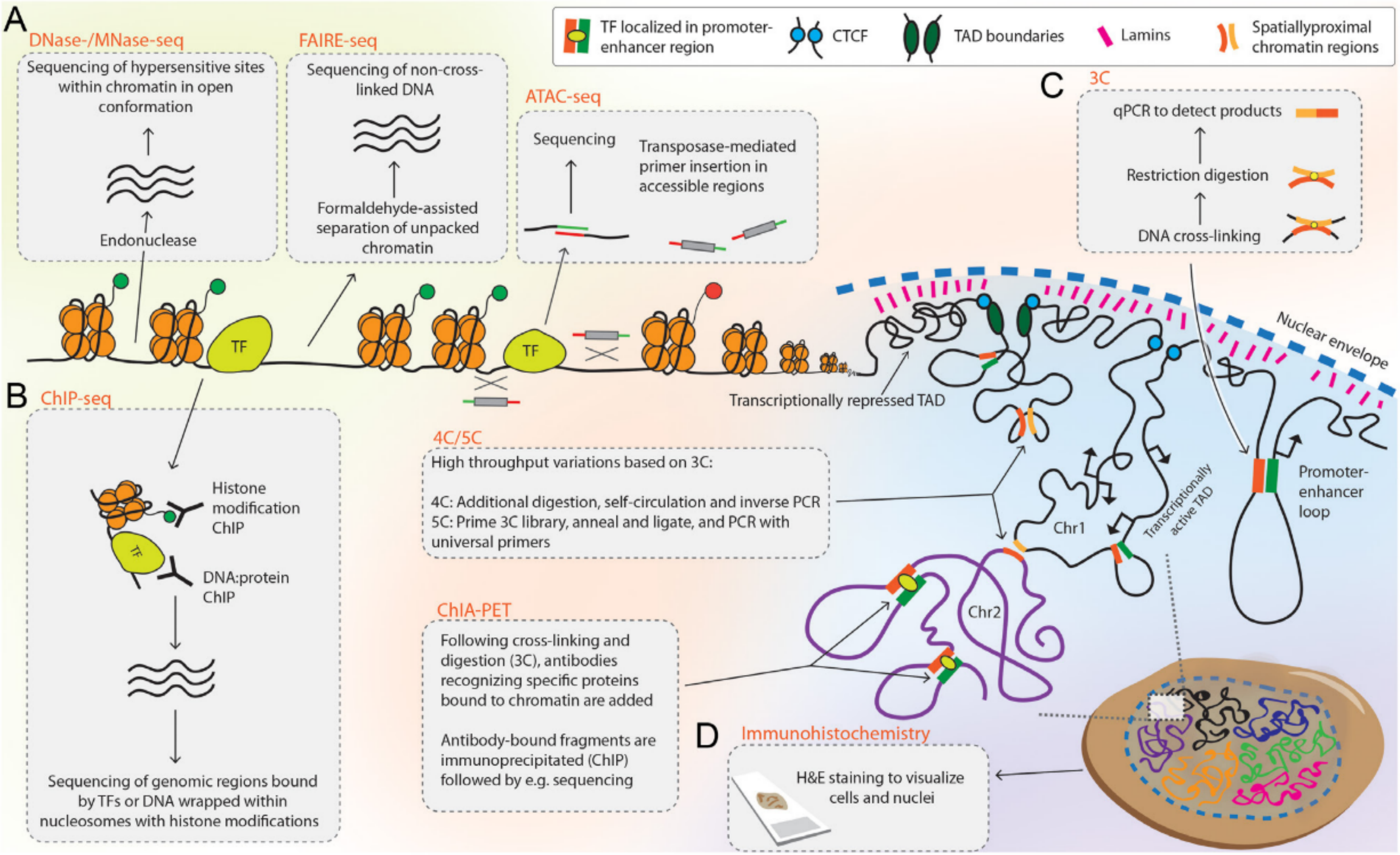
Nature reviews Genetics, 2004

Chromatin and Chromatin immunoprecipitation


Histones
Antibody
Specific Transcription Factor
Histone acetyltransferase
General Transcription Factors
RNA pol II
Active Gene
Cross-linkable complex

Endocrine-Related Cancer, 2018

Sequencing Application in Epigenome



A DNase-/MNase-seq

Sequencing of hypersensitive sites within chromatin in open conformation

Endonuclease

FAIRE-seq

Sequencing of non-cross-linked DNA

Formaldehyde-assisted separation of unpacked chromatin

ATAC-seq

Sequencing

Transposase-mediated primer insertion in accessible regions

B ChIP-seq

Histone modification ChIP

DNA:protein ChIP

Sequencing of genomic regions bound by TFs or DNA wrapped within nucleosomes with histone modifications

C 3C

qPCR to detect products

Restriction digestion

DNA cross-linking

4C/5C

High throughput variations based on 3C:

4C: Additional digestion, self-circulation and inverse PCR
5C: Prime 3C library, anneal and ligate, and PCR with universal primers

ChIA-PET

Following cross-linking and digestion (3C), antibodies recognizing specific proteins bound to chromatin are added

Antibody-bound fragments are immunoprecipitated (ChIP) followed by e.g. sequencing

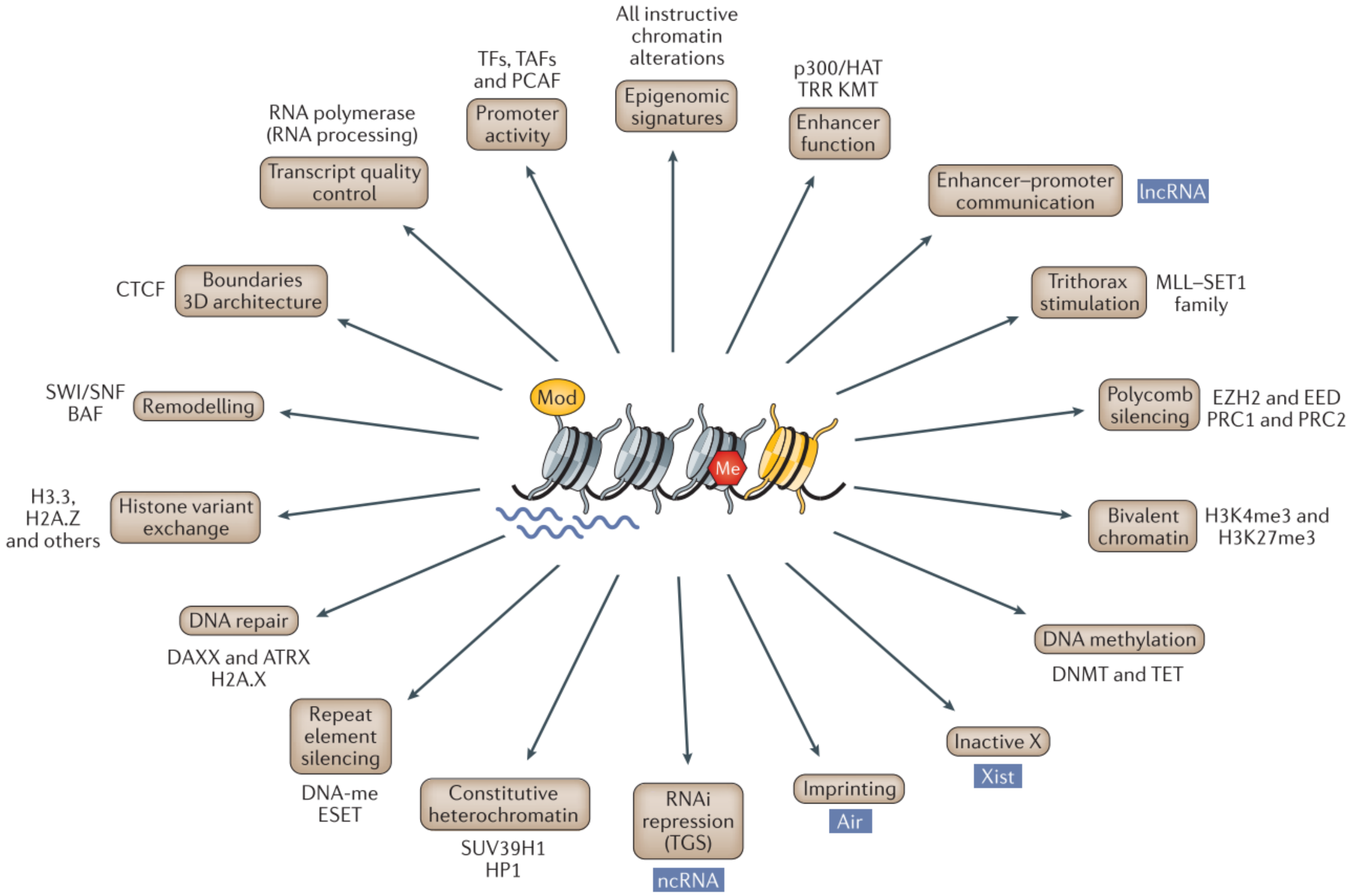
D Immunohistochemistry

H&E staining to visualize cells and nuclei

TF localized in promoter-enhancer region
CTCF
TAD boundaries
Lamins
Spatiallyproximal chromatin regions
TF
TF
TF
Transcriptionally repressed TAD
Nuclear envelope
Transcriptionally active TAD
Chr1
Chr2
Promoter-enhancer loop

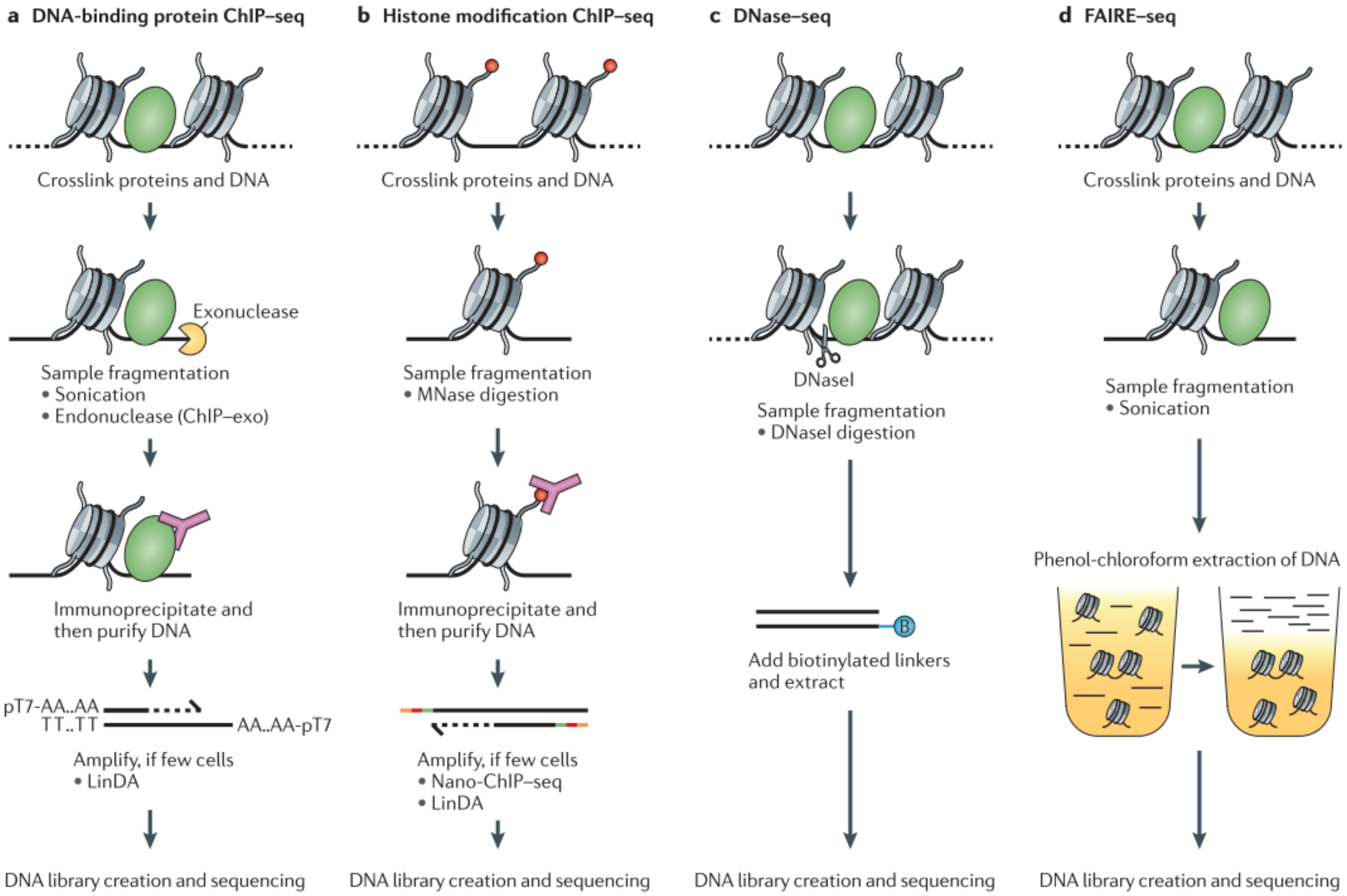
Endocrine-Related Cancer, 2018

Epigenome function



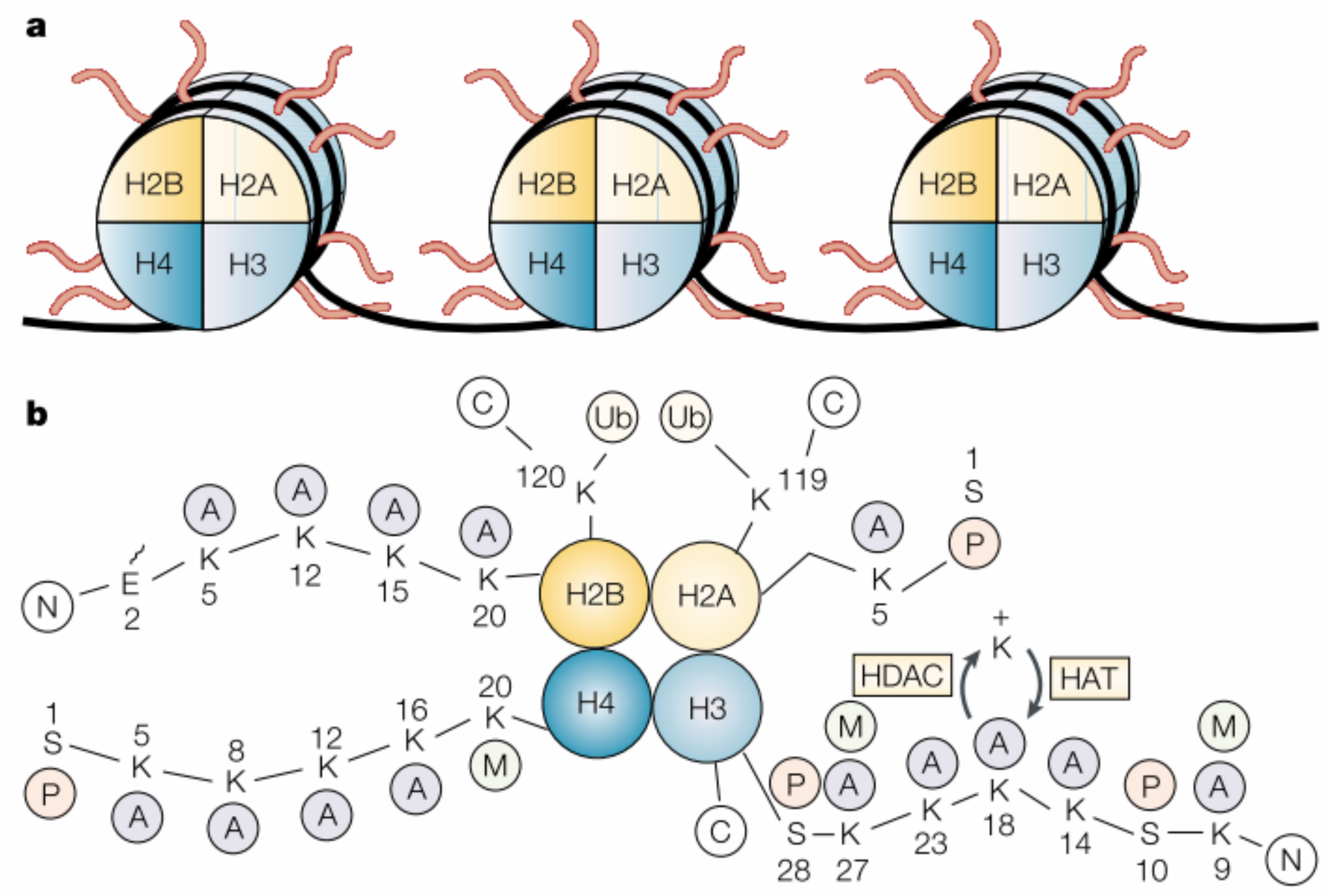
Nature Reviews Genetics, 2016

ChIP-Seq and DNA accessibility sequencing


a DNA-binding protein ChIP–seq
Crosslink proteins and DNA
Exonuclease
Sample fragmentation
• Sonication
• Endonuclease (ChIP–exo)
Immunoprecipitate and then purify DNA
pT7-AA..AA
TT..TT
AA..AA-pT7
Amplify, if few cells
• LinDA
DNA library creation and sequencing
b Histone modification ChIP–seq
Crosslink proteins and DNA
Sample fragmentation
• MNase digestion
Immunoprecipitate and then purify DNA
Amplify, if few cells
• Nano-ChIP–seq
• LinDA
DNA library creation and sequencing
c DNase–seq
DNaseI
Sample fragmentation
• DNaseI digestion
B
Add biotinylated linkers and extract
DNA library creation and sequencing
d FAIRE–seq
Crosslink proteins and DNA
Sample fragmentation
• Sonication
Phenol-chloroform extraction of DNA
DNA library creation and sequencing

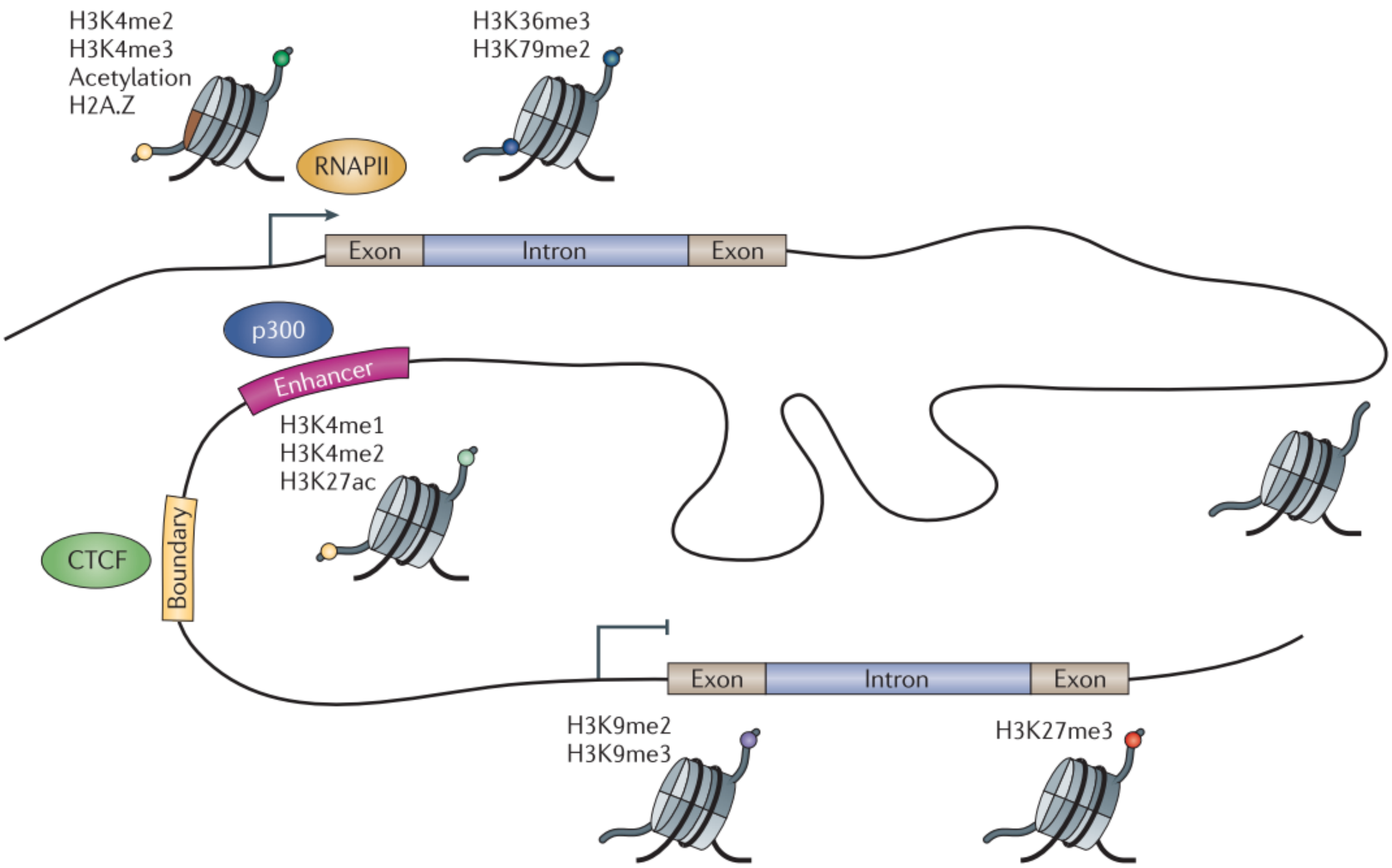
Nature Reviews Genetics, 2012

Histone modification



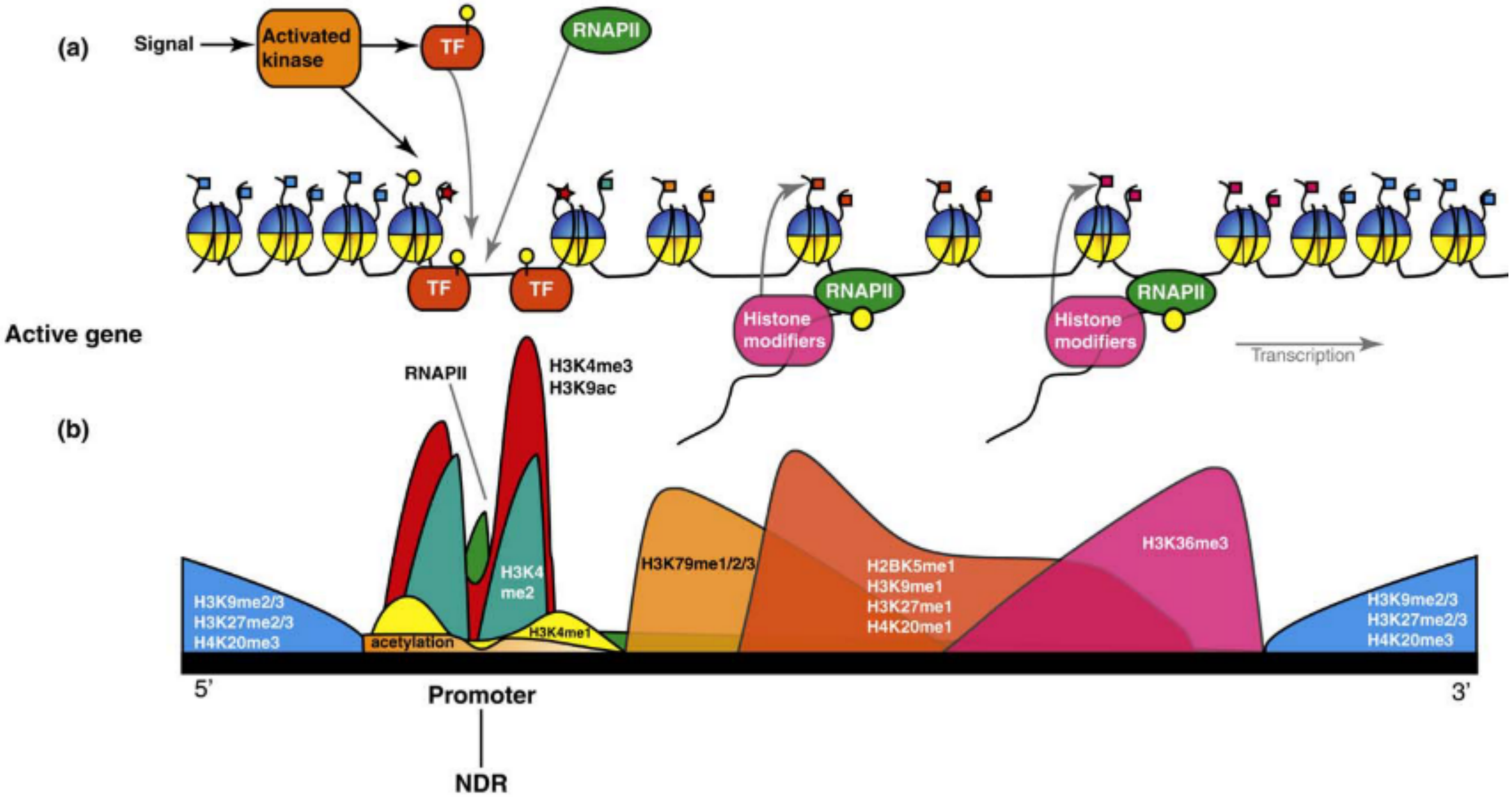
Nature reviews Cancer, 2001

Regulation by histone profile



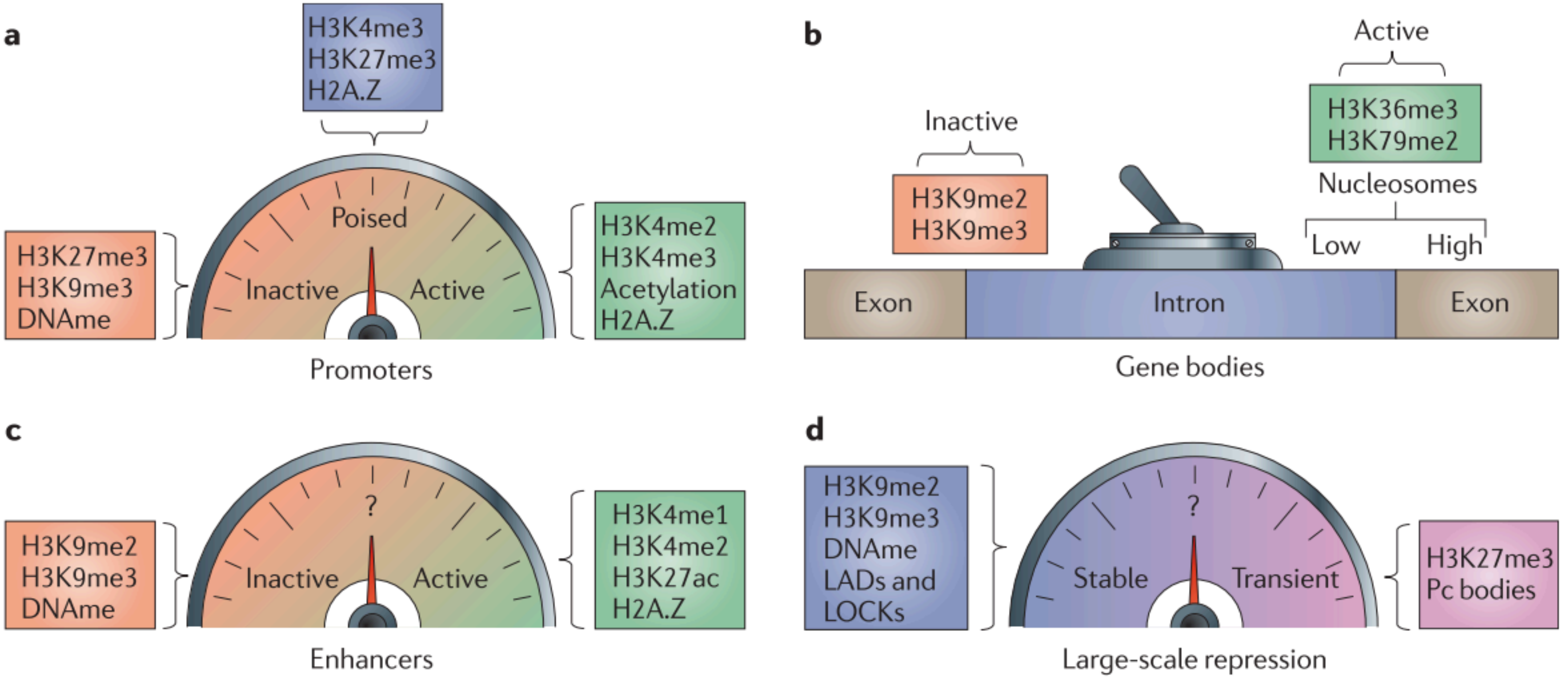
Nature reviews. Genetics, 2011

Active regulation in gene



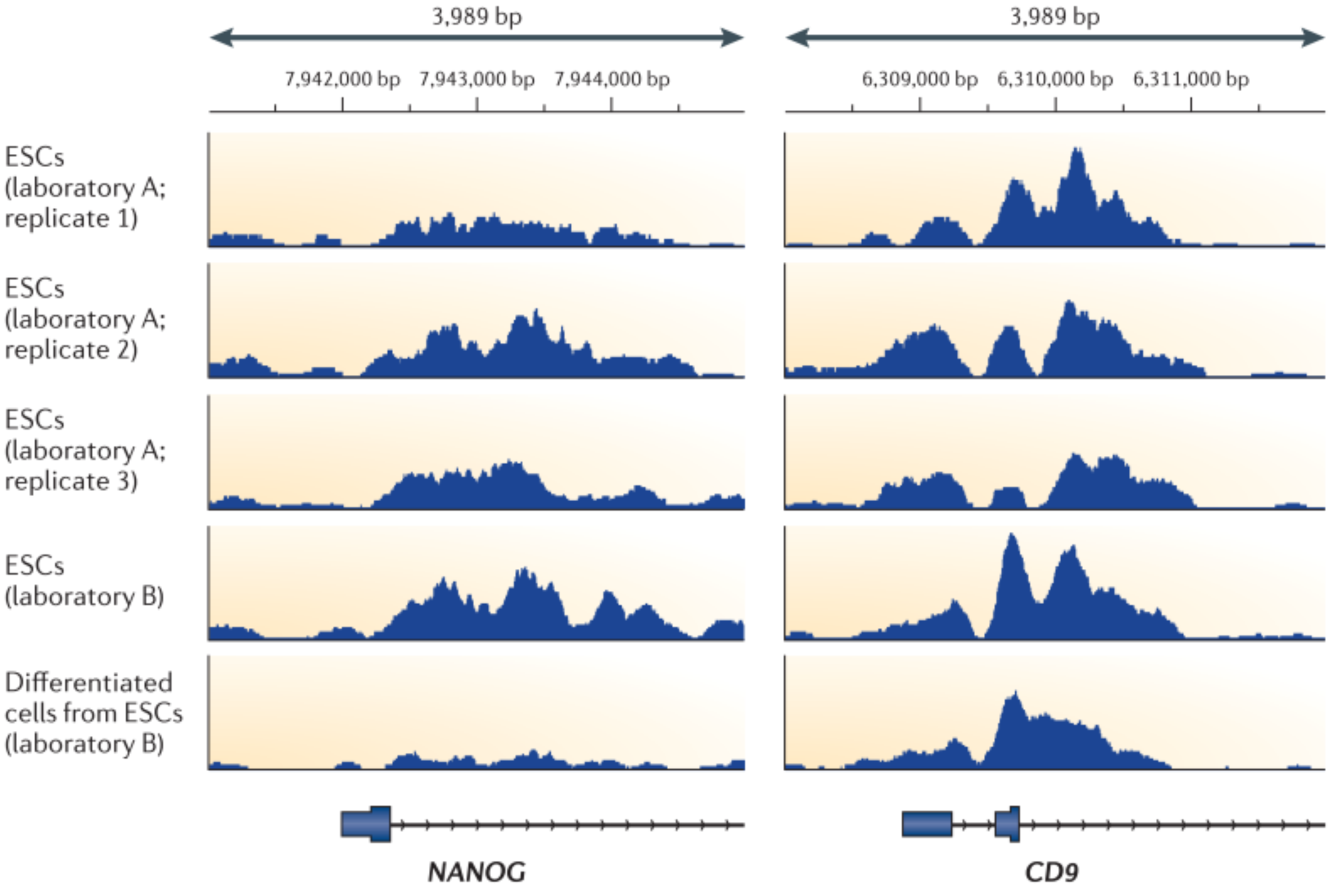
Trends in Biochemical Sciences, 2010

'Dashboard' of histone modifications



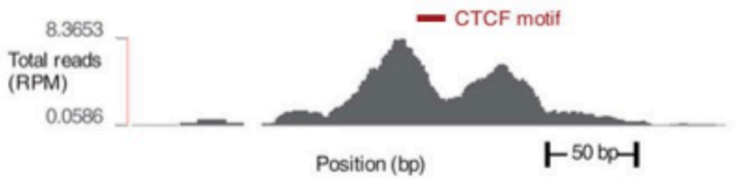
Nature reviews. Genetics, 2011

Histone profile depending on Cell type



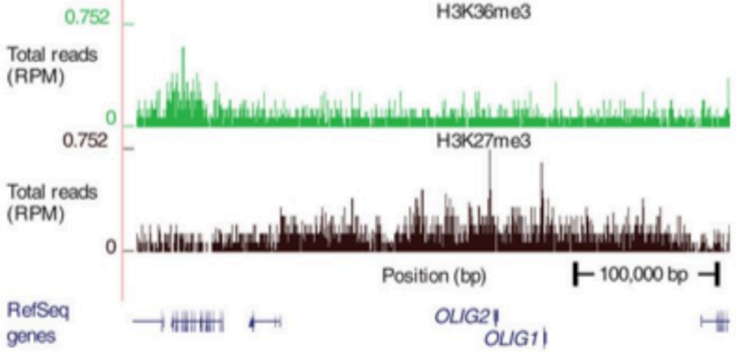
***Nature Reviews Genetics*, 2014**

Classes of DNA-bound proteins



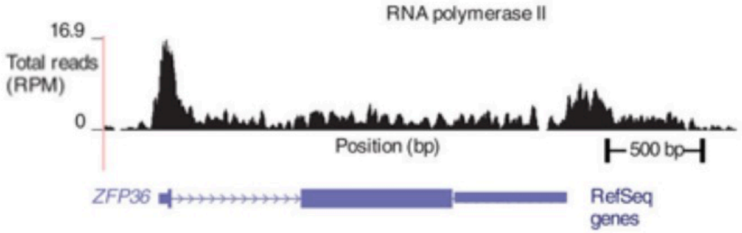
Point-source factors

Peak
(100bp ~ 1kbp)



Broad-source factors

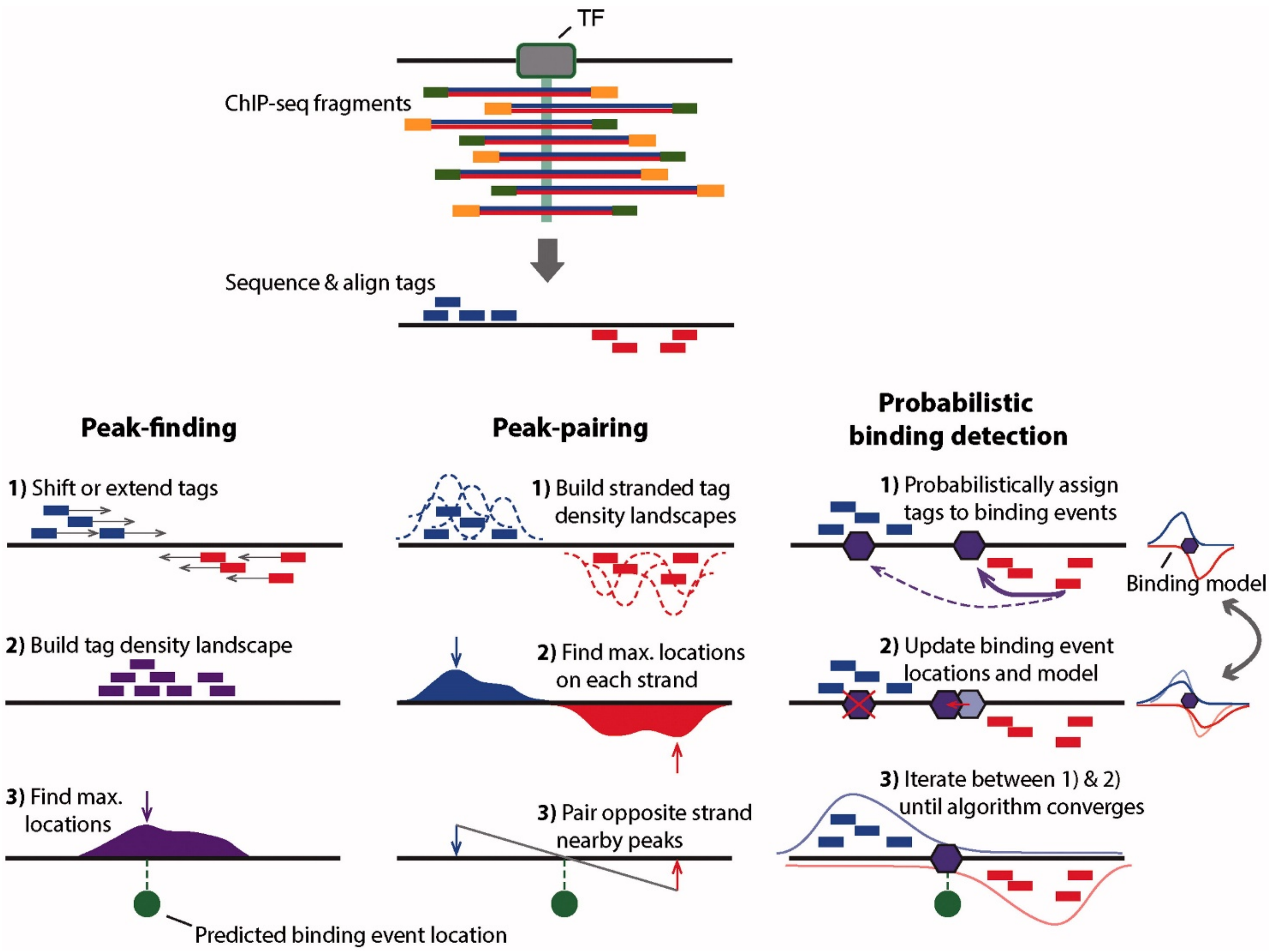
Domain
(kbp ~ Mbp)



Mixed-source factors

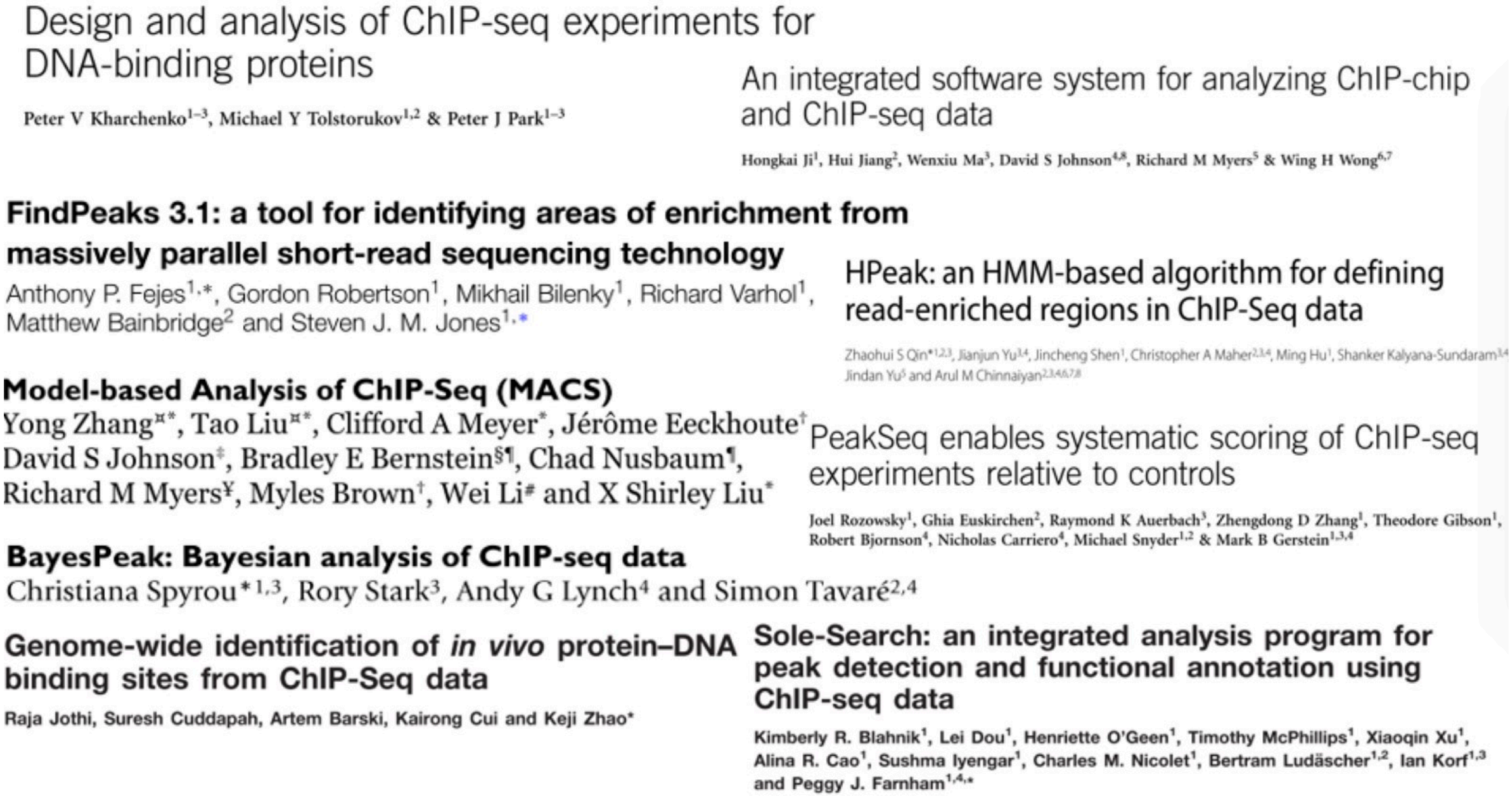
Nature Reviews Genetics, 2014

Principle detection of protein binding region


TF
ChIP-seq fragments
Sequence & align tags
Peak-finding
Peak-pairing
Probabilistic binding detection
1) Shift or extend tags
1) Build stranded tag density landscapes
1) Probabilistically assign tags to binding events
Binding model
2) Build tag density landscape
2) Find max. locations on each strand
2) Update binding event locations and model
3) Find max. locations
3) Pair opposite strand nearby peaks
3) Iterate between 1) & 2) until algorithm converges
Predicted binding event location

https://galaxyproject.org/tutorials/chip/

Principle detection of protein binding region



Design and analysis of ChIP-seq experiments for DNA-binding proteins

Peter V Kharchenko[1–3], Michael Y Tolstorukov[1,2] & Peter J Park[1–3]

An integrated software system for analyzing ChIP-chip and ChIP-seq data

Hongkai Ji[1], Hui Jiang[2], Wenxiu Ma[3], David S Johnson[4,8], Richard M Myers[5] & Wing H Wong[6,7]

FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology

Anthony P. Fejes[1,*], Gordon Robertson[1], Mikhail Bilenky[1], Richard Varhol[1], Matthew Bainbridge[2] and Steven J. M. Jones[1,*]

HPeak: an HMM-based algorithm for defining read-enriched regions in ChIP-Seq data

Zhaohui S Qin*[1,2,3], Jianjun Yu[3,4], Jincheng Shen[1], Christopher A Maher[2,3,4], Ming Hu[1], Shanker Kalyana-Sundaram[3,4] Jindan Yu[5] and Arul M Chinnaiyan[2,3,4,6,7,8]

Model-based Analysis of ChIP-Seq (MACS)

Yong Zhang[¤*], Tao Liu[¤*], Clifford A Meyer[*], Jérôme Eeckhoute[†], David S Johnson[‡], Bradley E Bernstein[§¶], Chad Nusbaum[¶], Richard M Myers[¥], Myles Brown[†], Wei Li[#] and X Shirley Liu[*]

PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls

Joel Rozowsky[1], Ghia Euskirchen[2], Raymond K Auerbach[3], Zhengdong D Zhang[1], Theodore Gibson[1], Robert Bjornson[4], Nicholas Carriero[4], Michael Snyder[1,2] & Mark B Gerstein[1,3,4]

BayesPeak: Bayesian analysis of ChIP-seq data

Christiana Spyrou*[1,3], Rory Stark[3], Andy G Lynch[4] and Simon Tavaré[2,4]

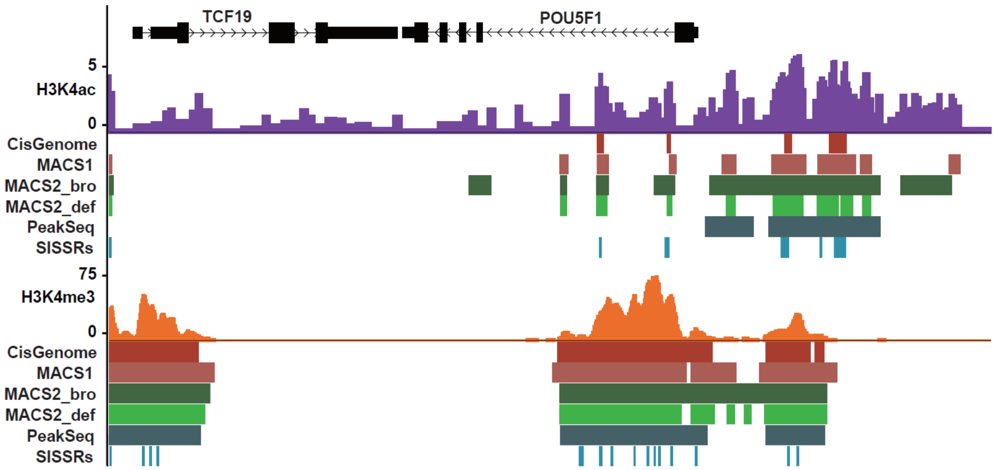
Genome-wide identification of *in vivo* protein–DNA binding sites from ChIP-Seq data

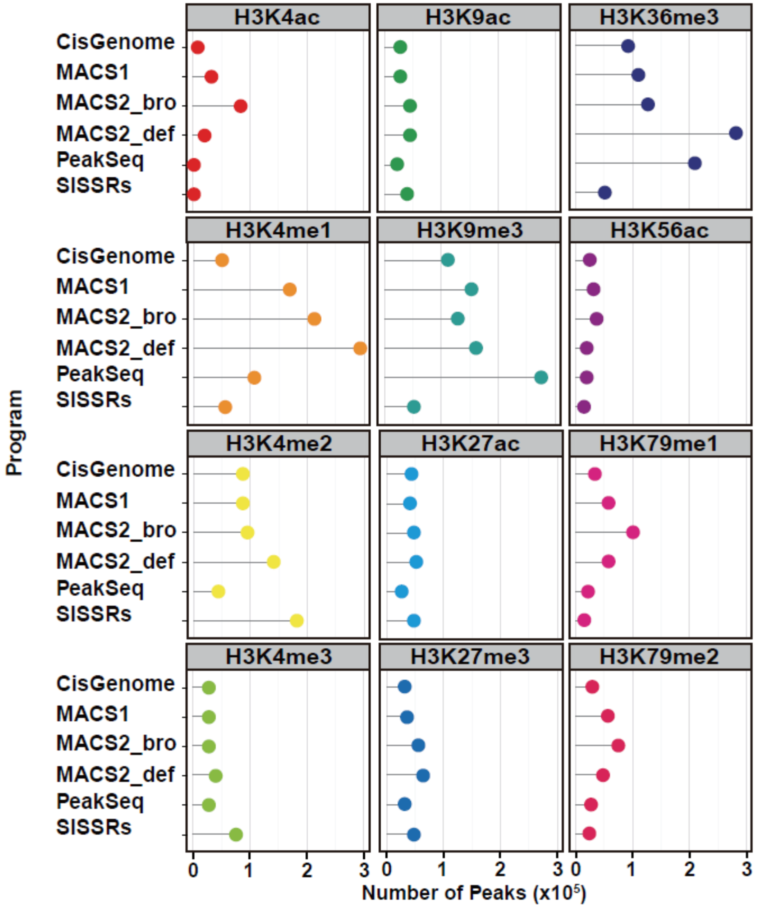
Raja Jothi, Suresh Cuddapah, Artem Barski, Kairong Cui and Keji Zhao*

Sole-Search: an integrated analysis program for peak detection and functional annotation using ChIP-seq data

Kimberly R. Blahnik[1], Lei Dou[1], Henriette O'Geen[1], Timothy McPhillips[1], Xiaoqin Xu[1], Alina R. Cao[1], Sushma Iyengar[1], Charles M. Nicolet[1], Bertram Ludäscher[1,2], Ian Korf[1,3] and Peggy J. Farnham[1,4,*]

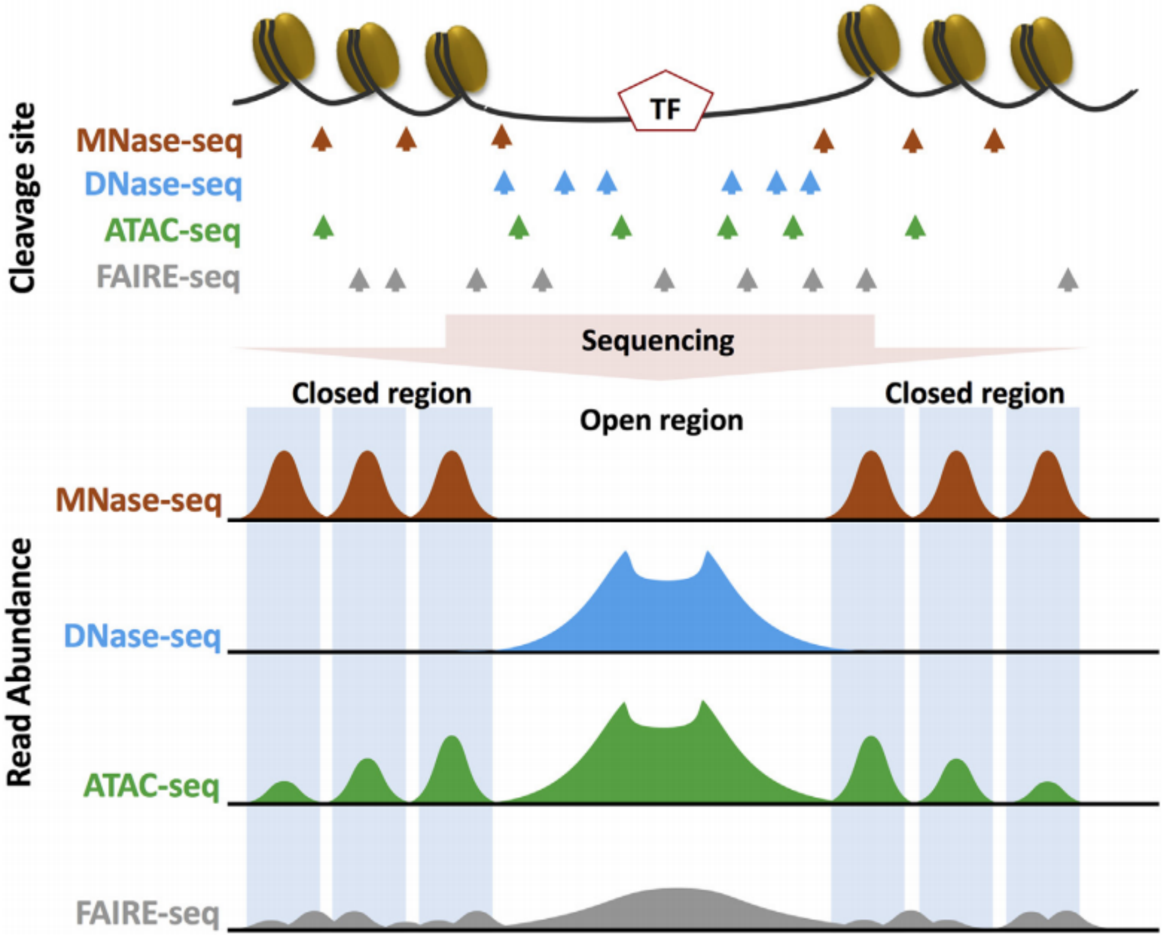
Difference of peak result depending on program





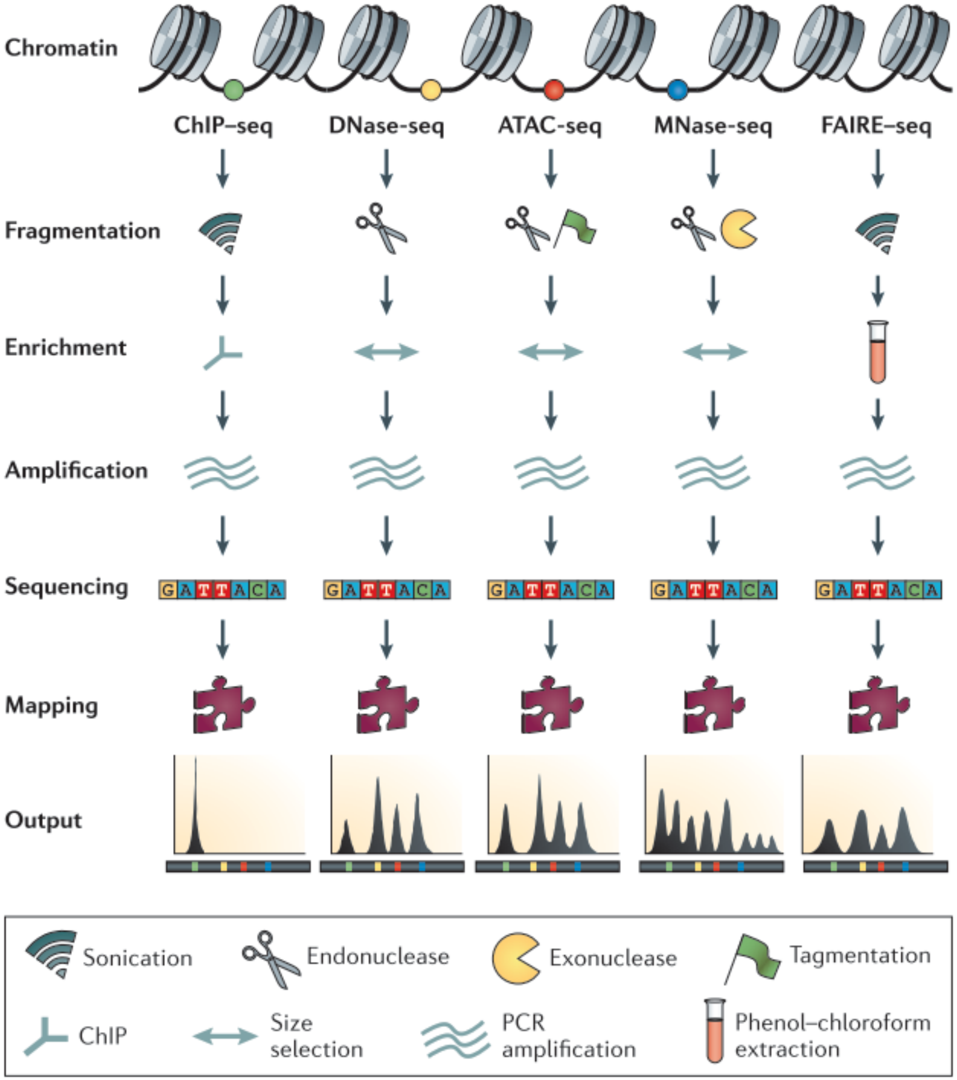
Unpublished

Active regulation in gene



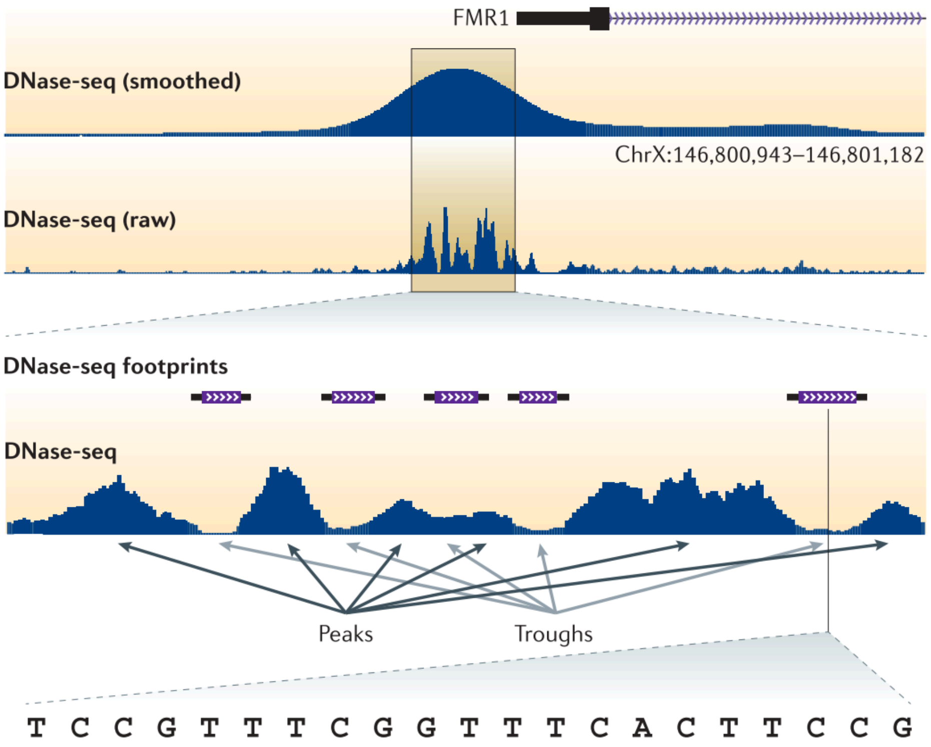
Epigenetics in Human Disease, 2018

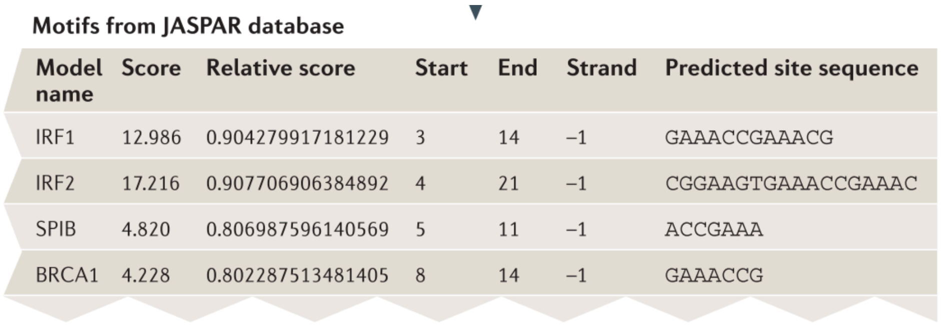
Sequencing to investigate DNA accessibility



Nature Reviews Genetics, 2014

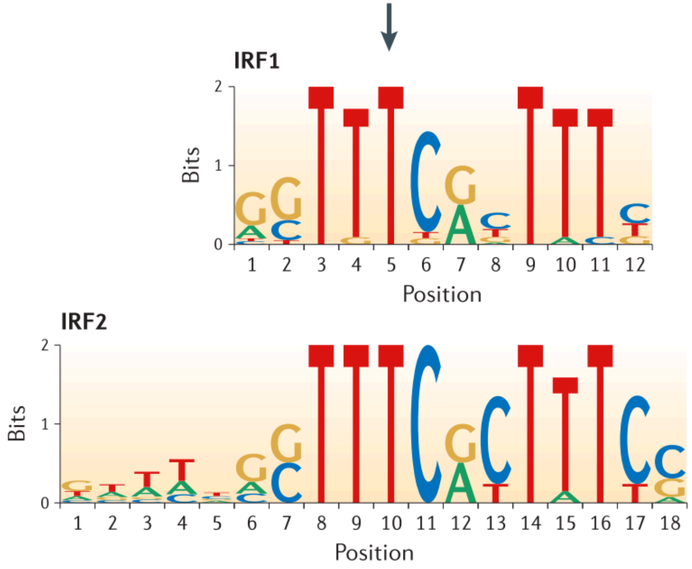
Sequencing to investigate DNA accessibility





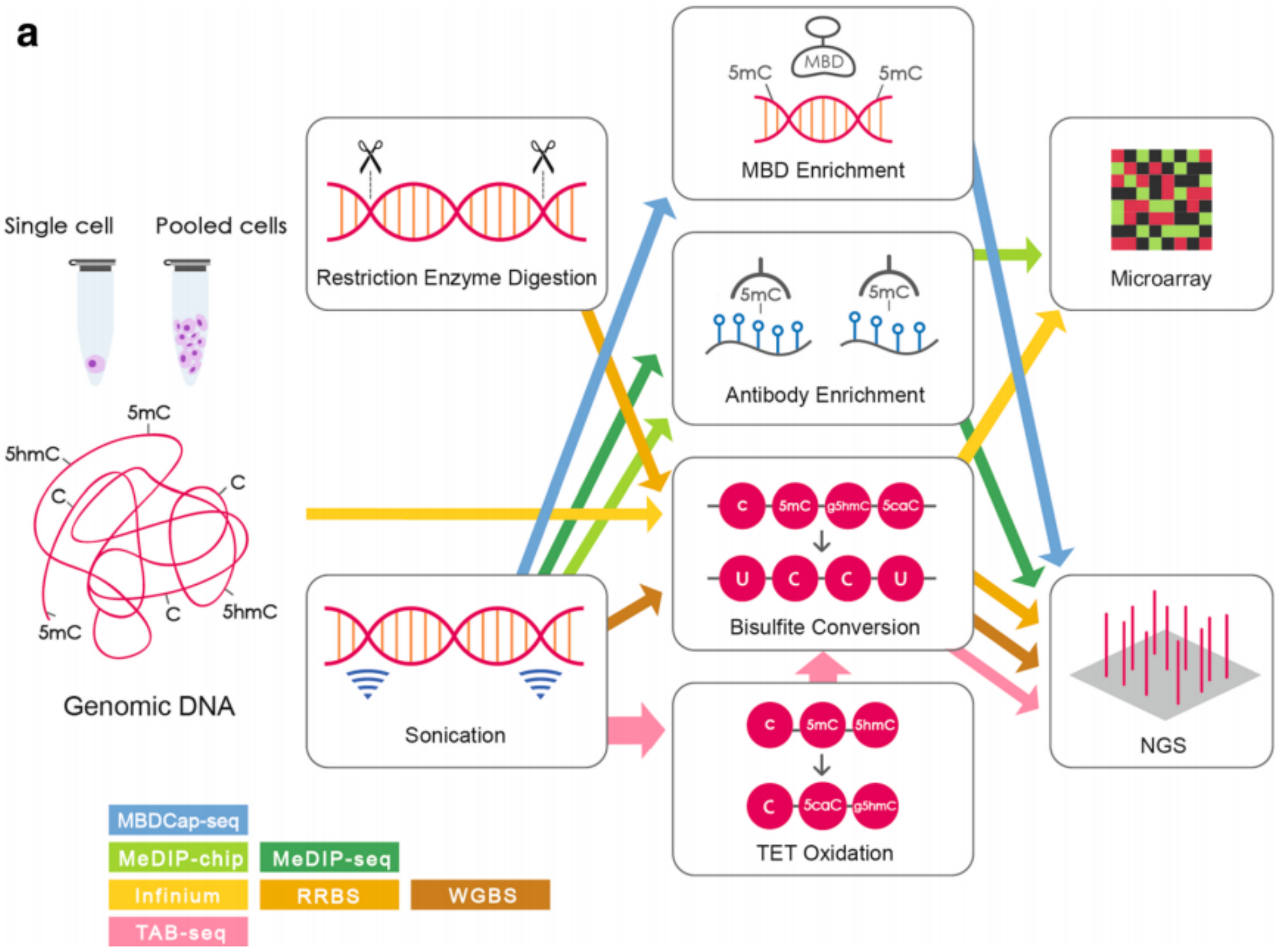
Motifs from JASPAR database

Model name	Score	Relative score	Start	End	Strand	Predicted site sequence
IRF1	12.986	0.904279917181229	3	14	−1	GAAACCGAAACG
IRF2	17.216	0.907706906384892	4	21	−1	CGGAAGTGAAACCGAAAC
SPIB	4.820	0.806987596140569	5	11	−1	ACCGAAA
BRCA1	4.228	0.802287513481405	8	14	−1	GAAACCG



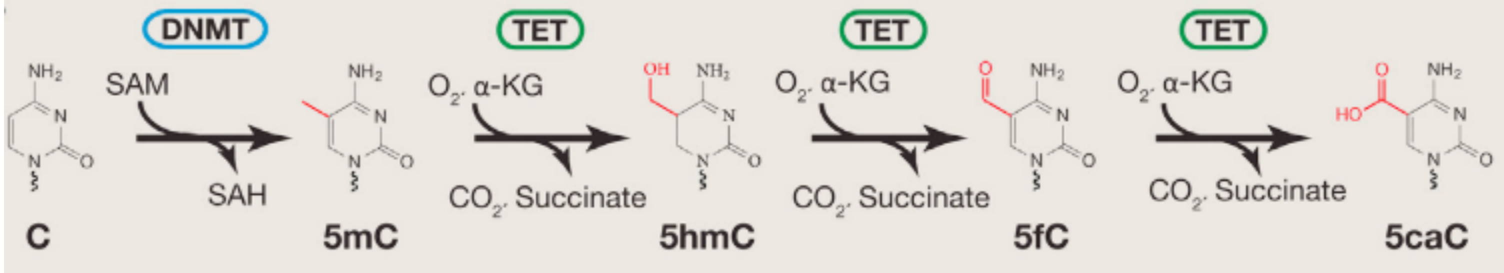
Nature Reviews Genetics, 2014

Sequencing method to measure methylation



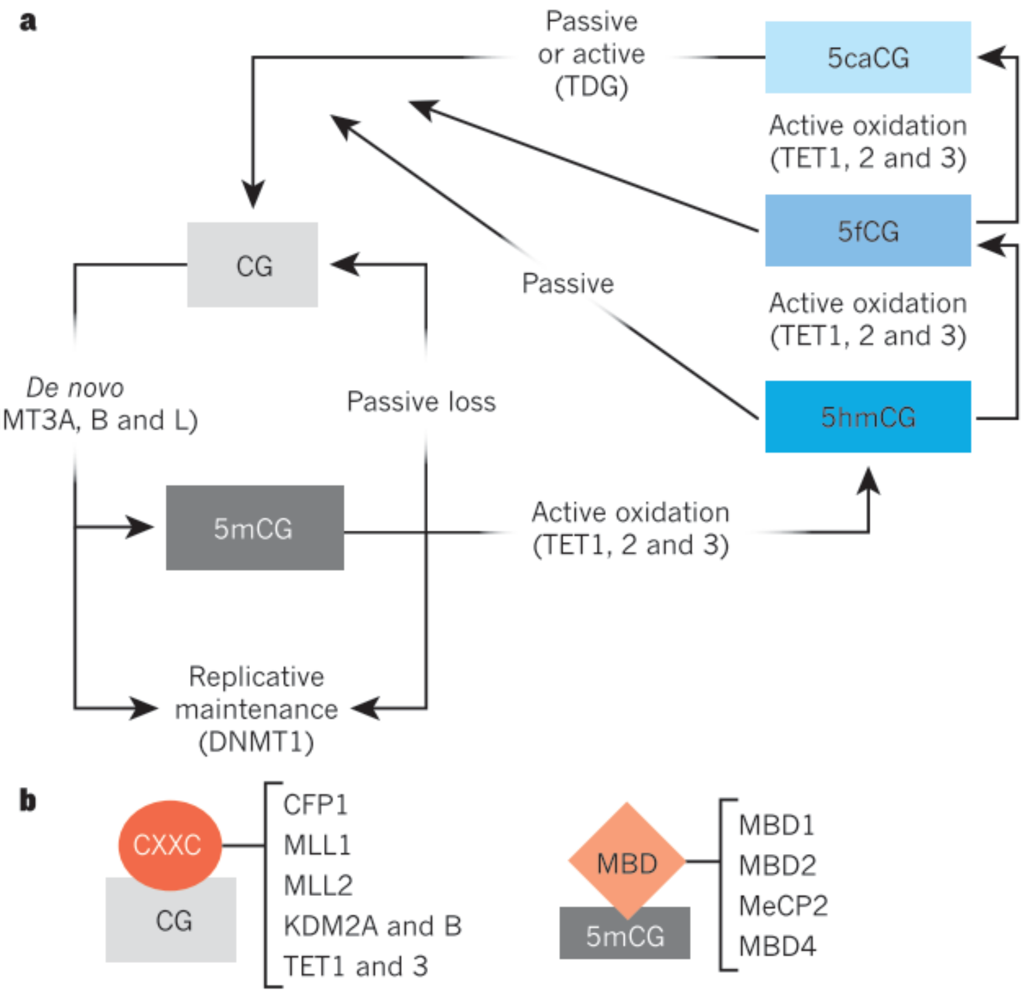
Cell, 2014

DNA methylation mechanism



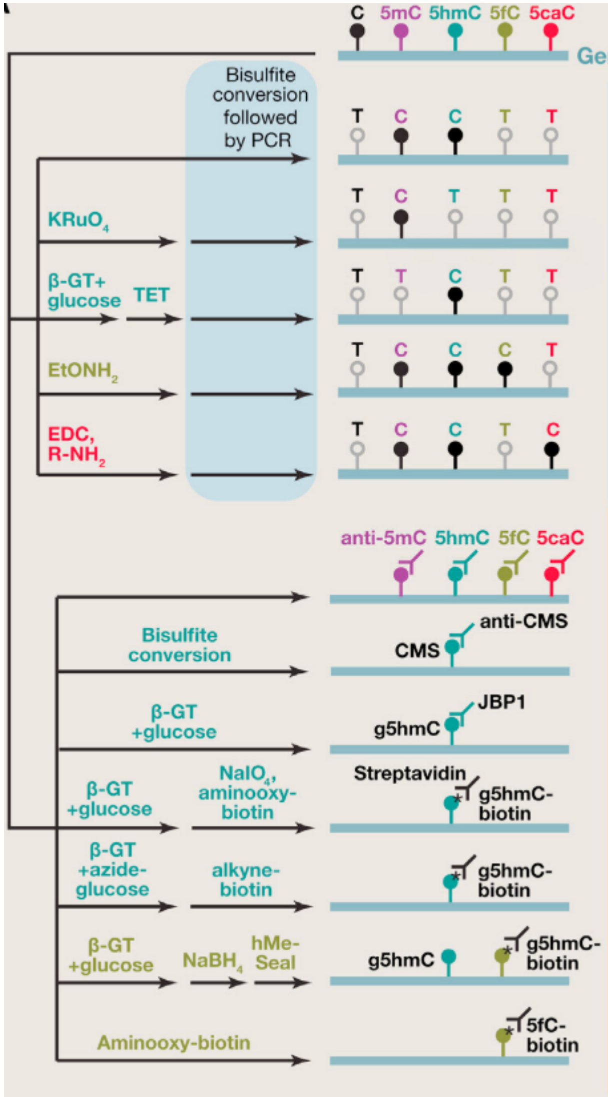
Cell, 2014

DNA methylation mechanism



Nature, 2015

Sequencing for methylation



Bisulfite sequencing based mapping methods

BS-Seq: directly map 5mC and 5hmC as C

oxBS-Seq: directly map 5mC as C, and map 5hmC after being subtracted from BS-Seq

TAB-Seq: directly map 5hmC as C

fCAB-Seq: map 5fC after subtracting BS-Seq

caCAB-Seq: map 5caC after subtracting BS-Seq

Affinity-enrichment based mapping methods

DIP: Map 5mC/5hmC/5fC/5caC using modification-specific antibodies

CMS: Map 5hmC using antibody specific for CMS

JBP1 IP: Map 5hmC using JBP1, a natural binding protein for g5hmC

GLIB: Map 5hmC by chemically tagging 5hmC with biotin

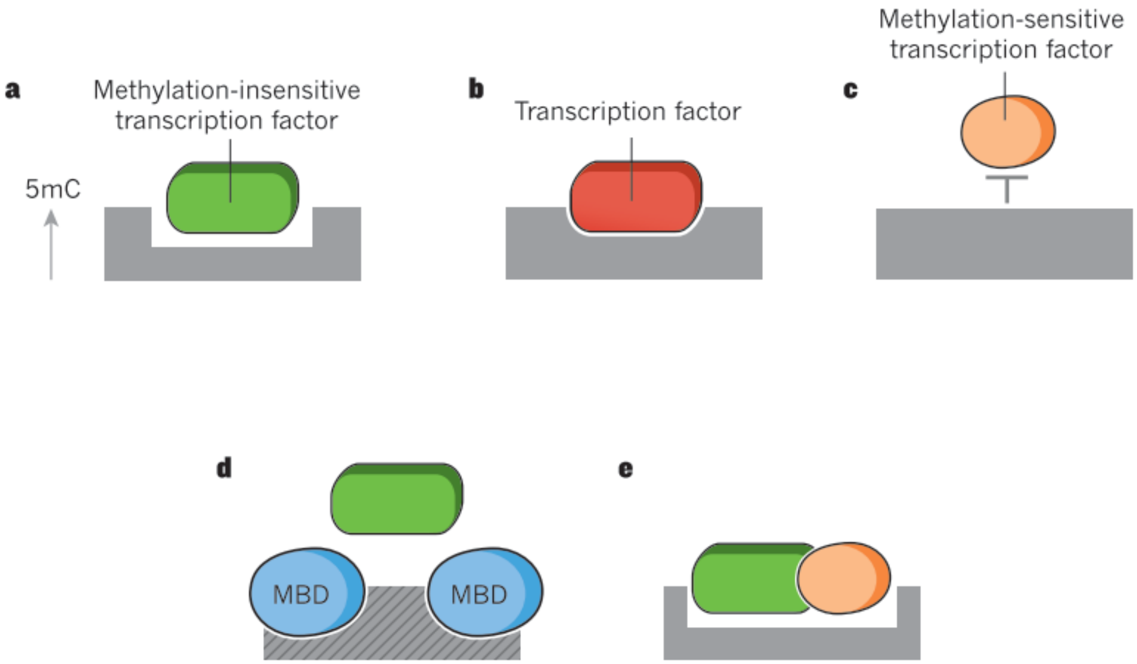
hMe-Seal: Map 5hmC by chemically tagging 5hmC with biotin

fC-Seal: Map 5fC by chemically tagging 5fC with biotin

5fC pull-down: Map 5fC by chemically tagging 5fC with biotin

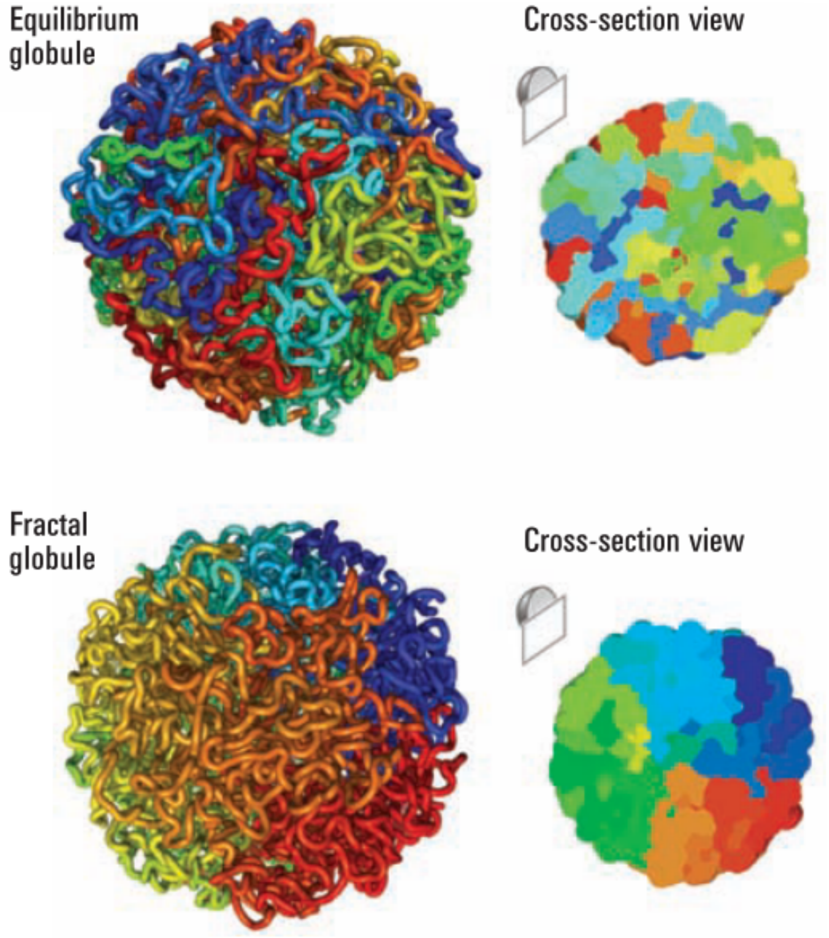
Cell, 2014

DNA methylation function


a
Methylation-insensitive transcription factor
5mC
b
Transcription factor
c
Methylation-sensitive transcription factor
d
MBD
MBD
e

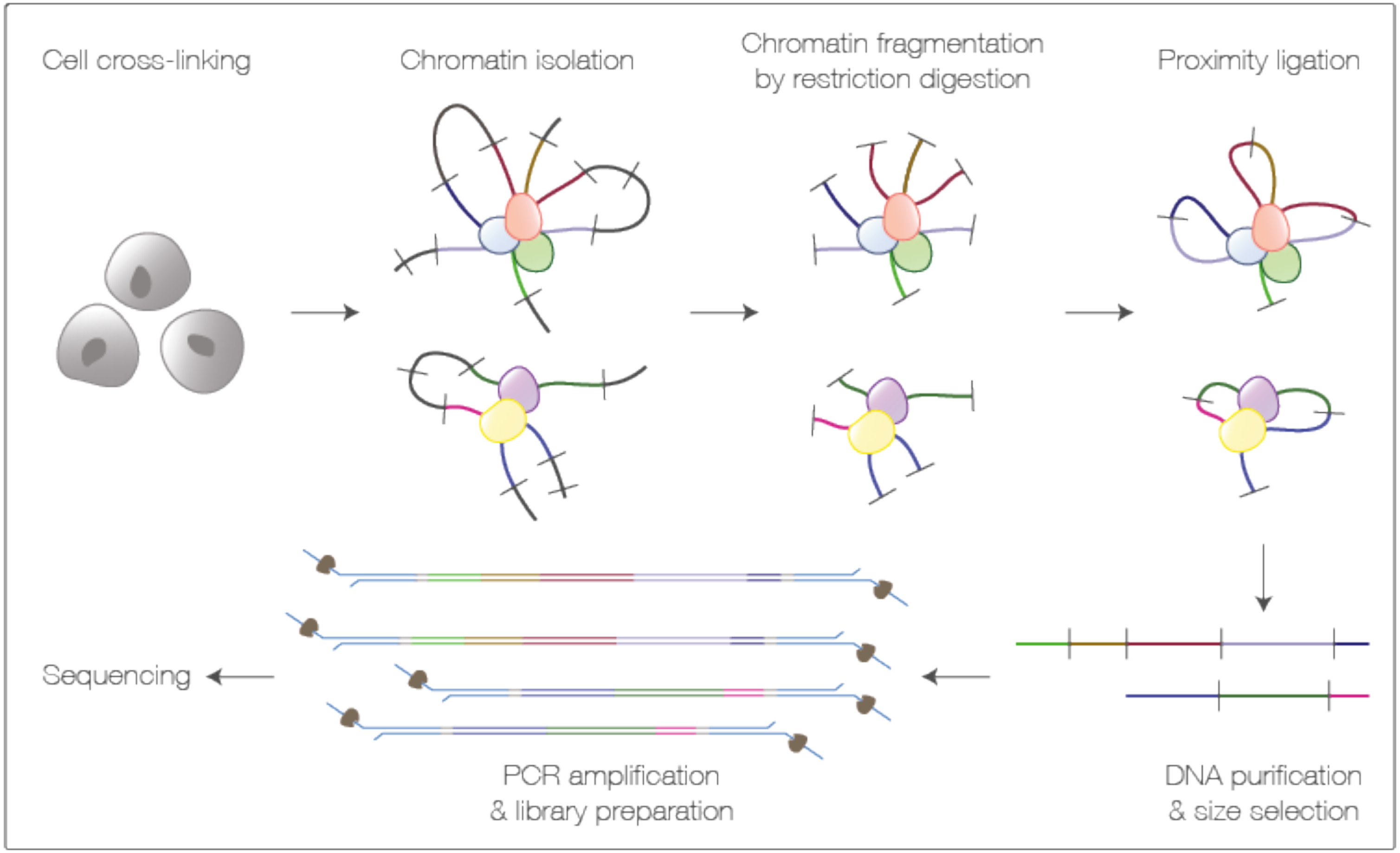
Nature, 2015

Chromatin structure



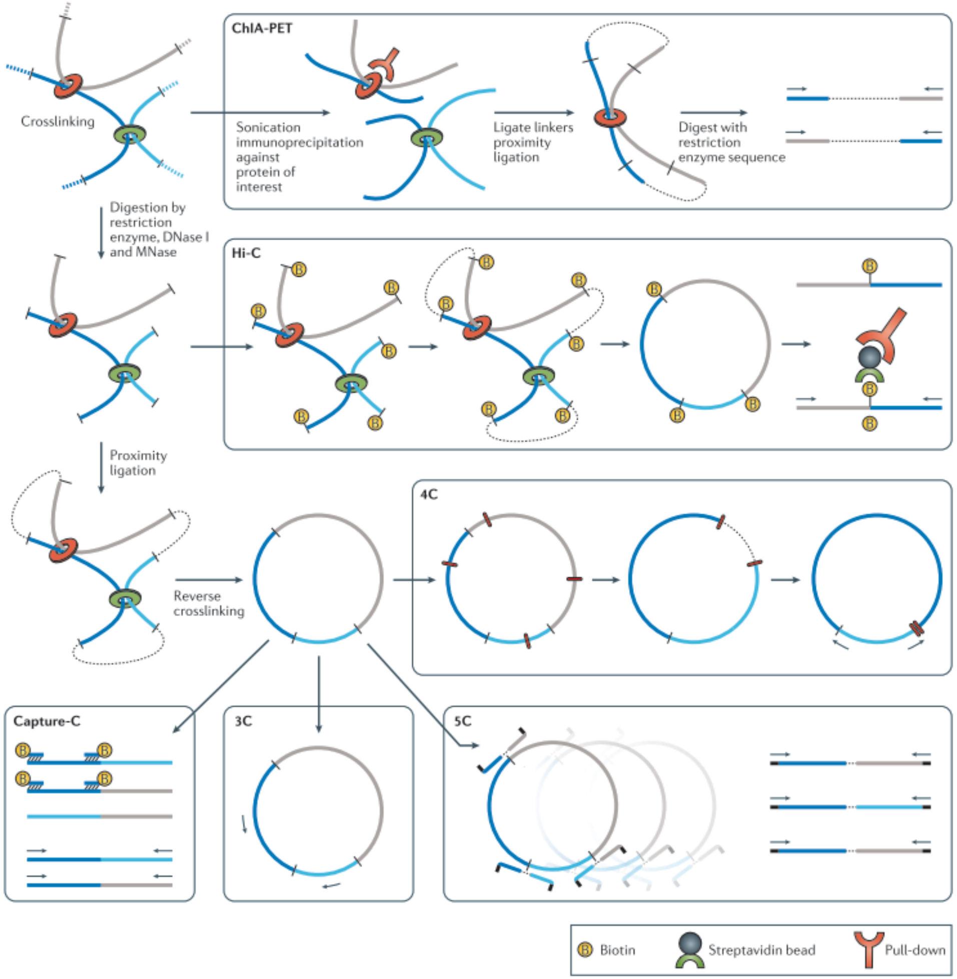
Trends in Biochemical Sciences, 2010

Chromatin Conformation Capture(3C)


Cell cross-linking
Chromatin isolation
Chromatin fragmentation by restriction digestion
Proximity ligation
DNA purification & size selection
PCR amplification & library preparation
Sequencing

https://nanoporetech.com/resource-centre/pore-c-using-nanopore-reads-delineate-long-range-interactions-between-genomic-loci#image1&

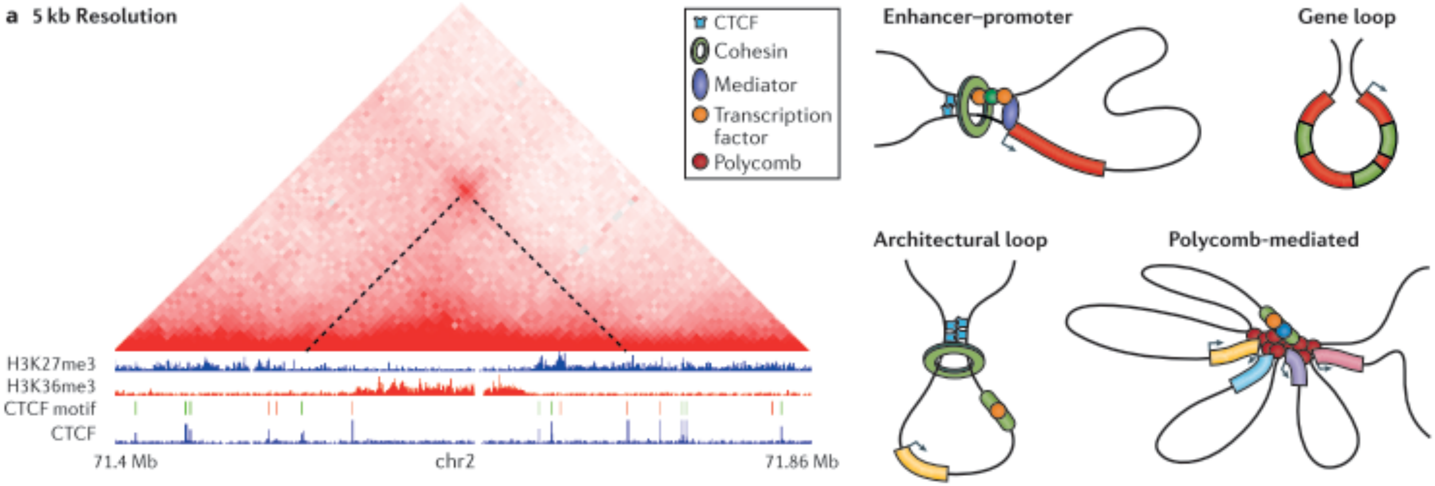
3C-based approaches



Crosslinking
Digestion by restriction enzyme, DNase I and MNase
Proximity ligation
Reverse crosslinking
ChIA-PET
Sonication immunoprecipitation against protein of interest
Ligate linkers proximity ligation
Digest with restriction enzyme sequence
Hi-C
B
4C
Capture-C
3C
5C
B Biotin
Streptavidin bead
Pull-down

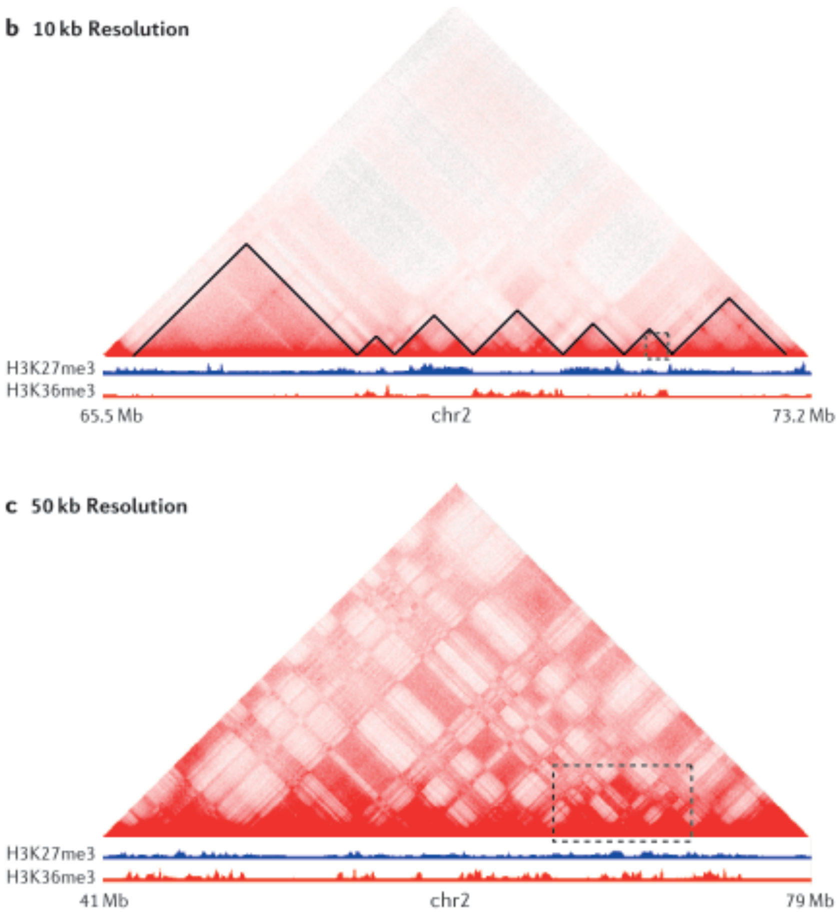
***Nature Reviews Genetics,* 2016**

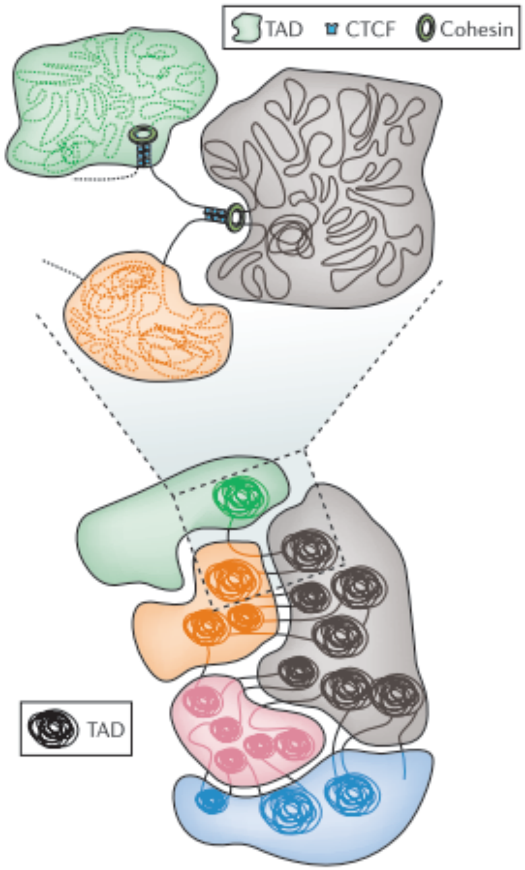
Conformation in genome



***Nature Reviews Genetics*, 2016**

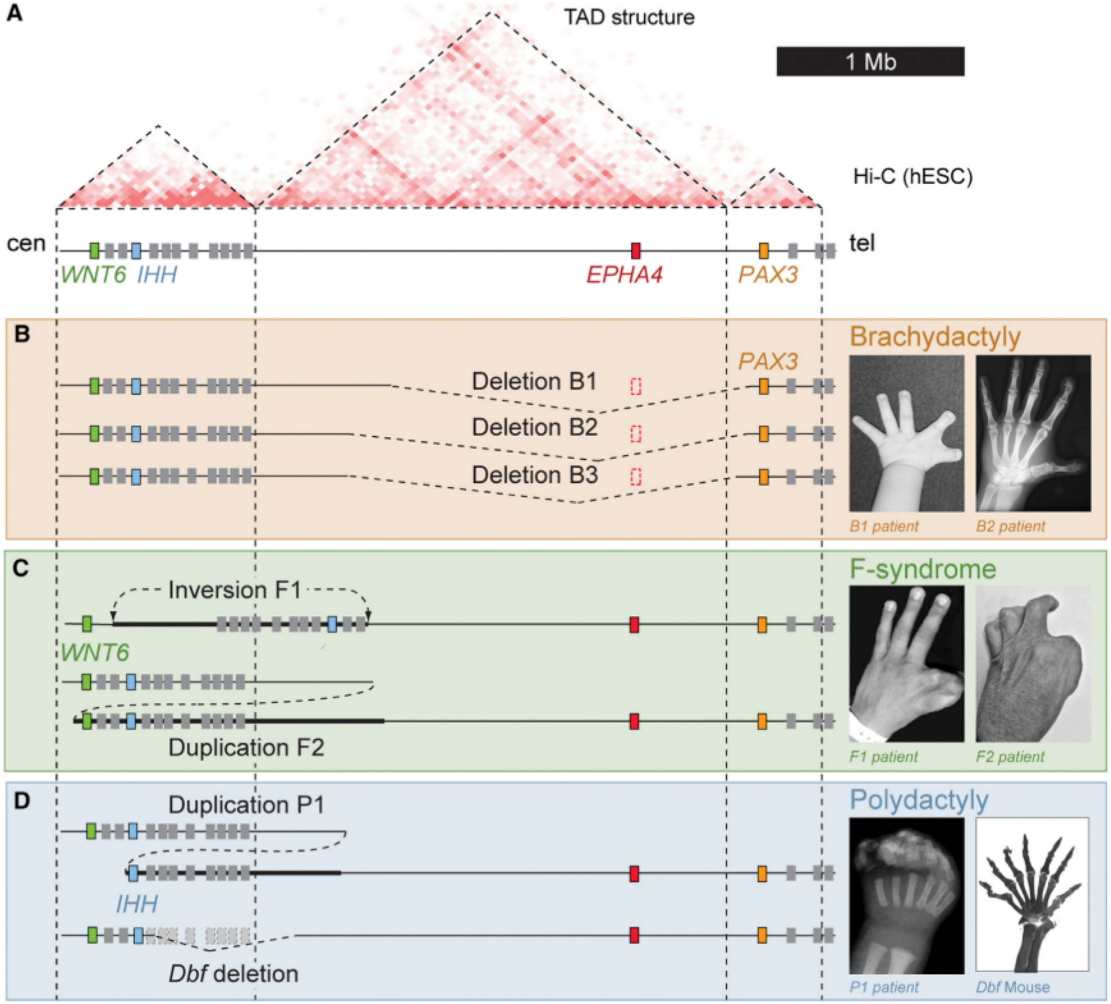
Conformation in genome





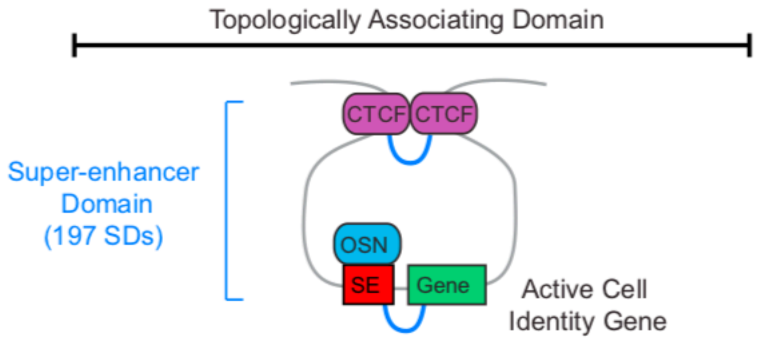
Nature Reviews Genetics, **2016**

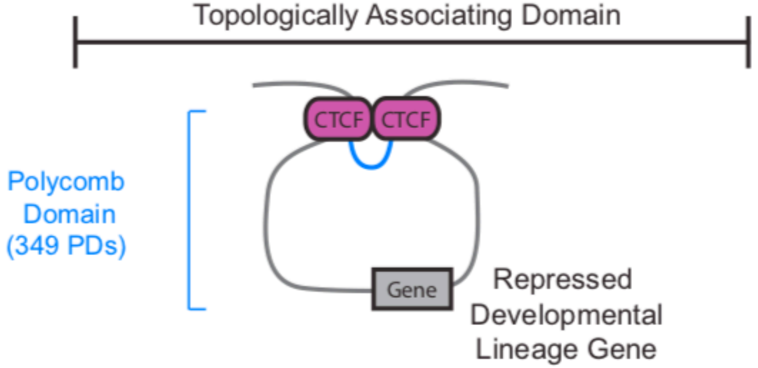
Diseases associated to chromatin conformation



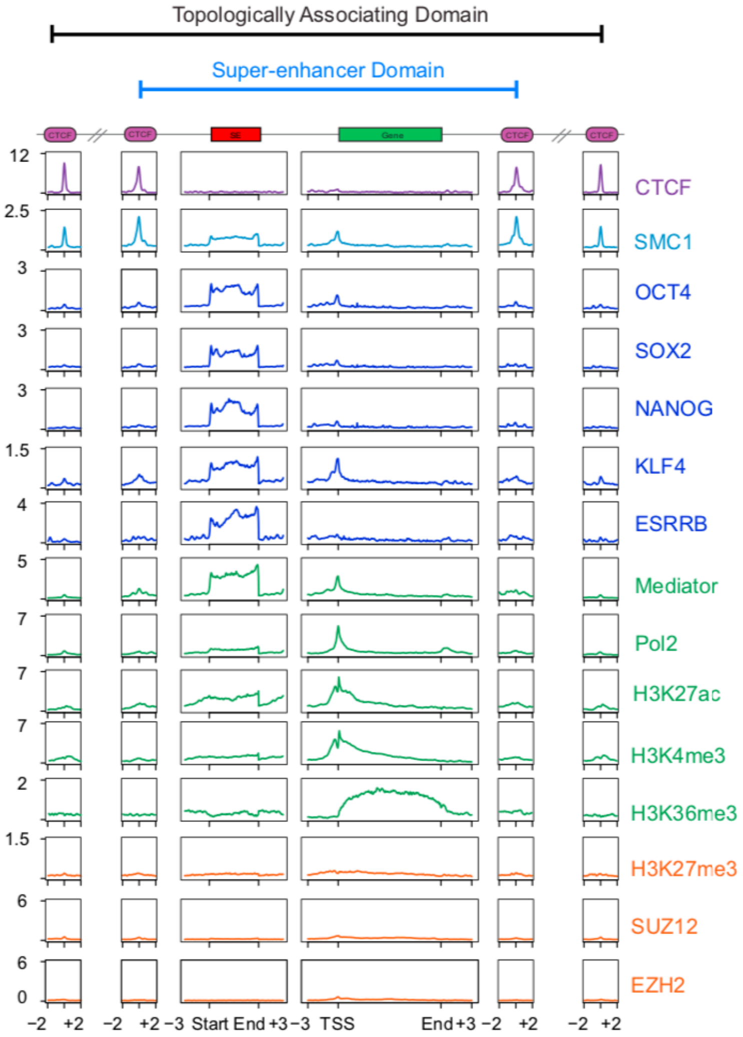
Cell, **2015**

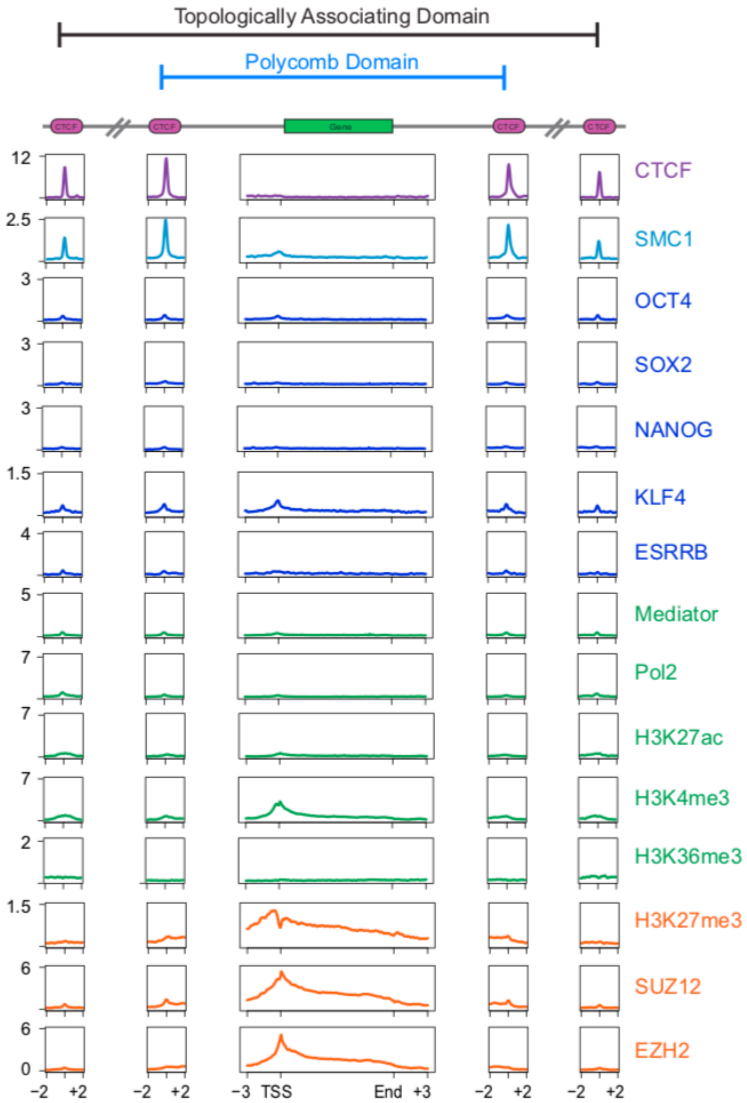
Interplay among epigenome feature


Topologically Associating Domain
Super-enhancer Domain (197 SDs)
CTCF
CTCF
OSN
SE
Gene
Active Cell Identity Gene

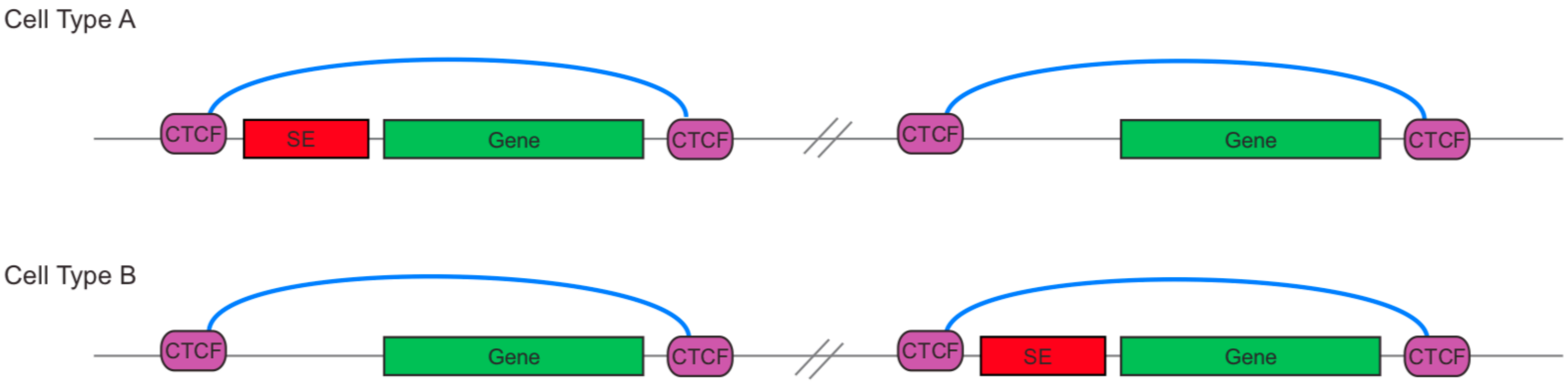

Topologically Associating Domain
Polycomb Domain (349 PDs)
CTCF
CTCF
Gene
Repressed Developmental Lineage Gene

Interplay among epigenome feature

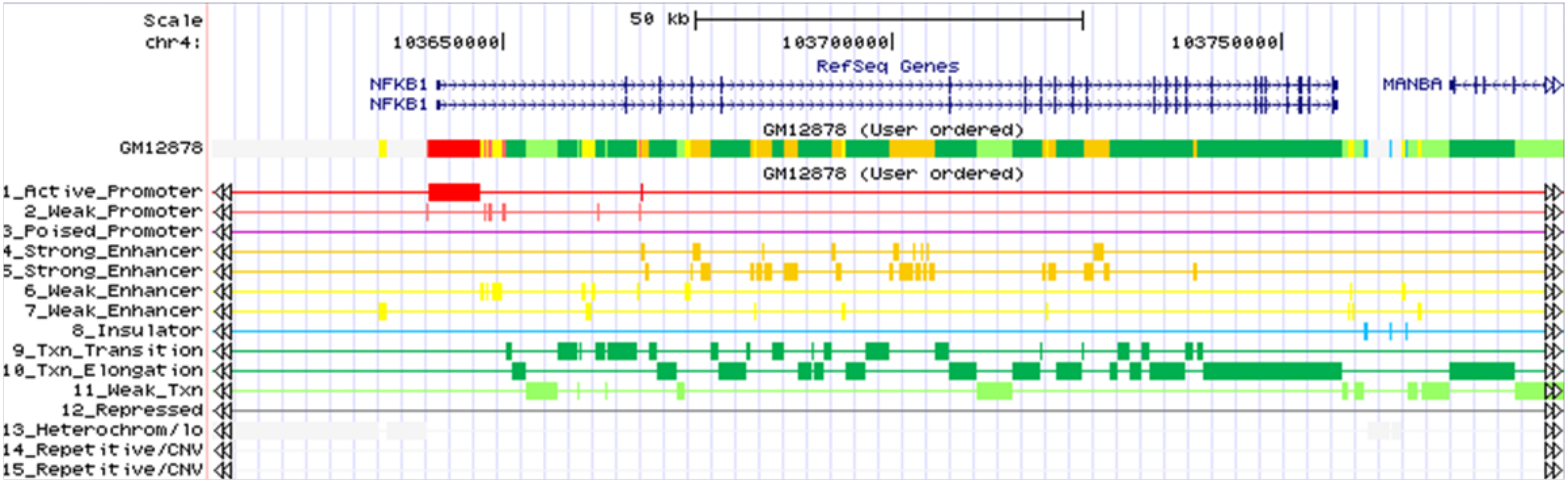




Interplay among epigenome feature



Multi-omics analysis



http://compbio.mit.edu/ChromHMM/

THANK YOU.

BMS